PreQuest User Guide

2010 CSDS Release

Copyright © 2009 The Cambridge Crystallographic Data Centre
Registered Charity No 800579

Conditions of Use

The Cambridge Structural Database System (CSD System) comprising all or some of the following: ConQuest, Quest, PreQuest, Mercury, (Mercury CSD and Materials module of Mercury), VISTA, Mogul, IsoStar, SuperStar, web accessible CSD tools and services, WebCSD, CSD Java sketcher, CSD data file, CSD-UNITY, CSD-MDL, CSD-SDfile, CSD data updates, sub files derived from the foregoing data files, documentation and command procedures (each individually a Component) is a database and copyright work belonging to the Cambridge Crystallographic Data Centre (CCDC) and its licensors and all rights are protected. Use of the CSD System is permitted solely in accordance with a valid Licence of Access Agreement and all Components included are proprietary. When a Component is supplied independently of the CSD System its use is subject to the conditions of the separate licence. All persons accessing the CSD System or its Components should make themselves aware of the conditions contained in the Licence of Access Agreement or the relevant licence.

In particular:

- The CSD System and its Components are licensed subject to a time limit for use by a specified organisation at a specified location.
- The CSD System and its Components are to be treated as confidential and may NOT be disclosed or re-distributed in any form, in whole or in part, to any third party.
- Software or data derived from or developed using the CSD System may not be distributed without prior written approval of the CCDC. Such prior approval is also needed for joint projects between academic and for-profit organisations involving use of the CSD System.
- The CSD System and its Components may be used for scientific research, including the design of novel compounds. Results may be published in the scientific literature, but each such publication must include an appropriate citation as indicated in the Schedule to the Licence of Access Agreement and on the CCDC website.
- No representations, warranties, or liabilities are expressed or implied in the supply of the CSD System or its Components by CCDC, its servants or agents, except where such exclusion or limitation is prohibited, void or unenforceable under governing law.

Licences may be obtained from:

Cambridge Crystallographic Data Centre
12 Union Road
Cambridge CB2 1EZ
United Kingdom

Web: *http://www.ccdc.cam.ac.uk*
Telephone: +44-1223-336408
Email: *admin@ccdc.cam.ac.uk*

(UNITY is a product of Tripos, L.P. and MDL is a registered trademark of Elsevier MDL)

1 INTRODUCTION

- PreQuest is a data conversion program provided as part of the Cambridge Structural Database System. The primary objective of the program is to create high quality structural data files in a format searchable by ConQuest.
- PreQuest enables users to build their own private databases of structures that are then searchable either independently of, or in unison with, the CSD.

2 GETTING STARTED

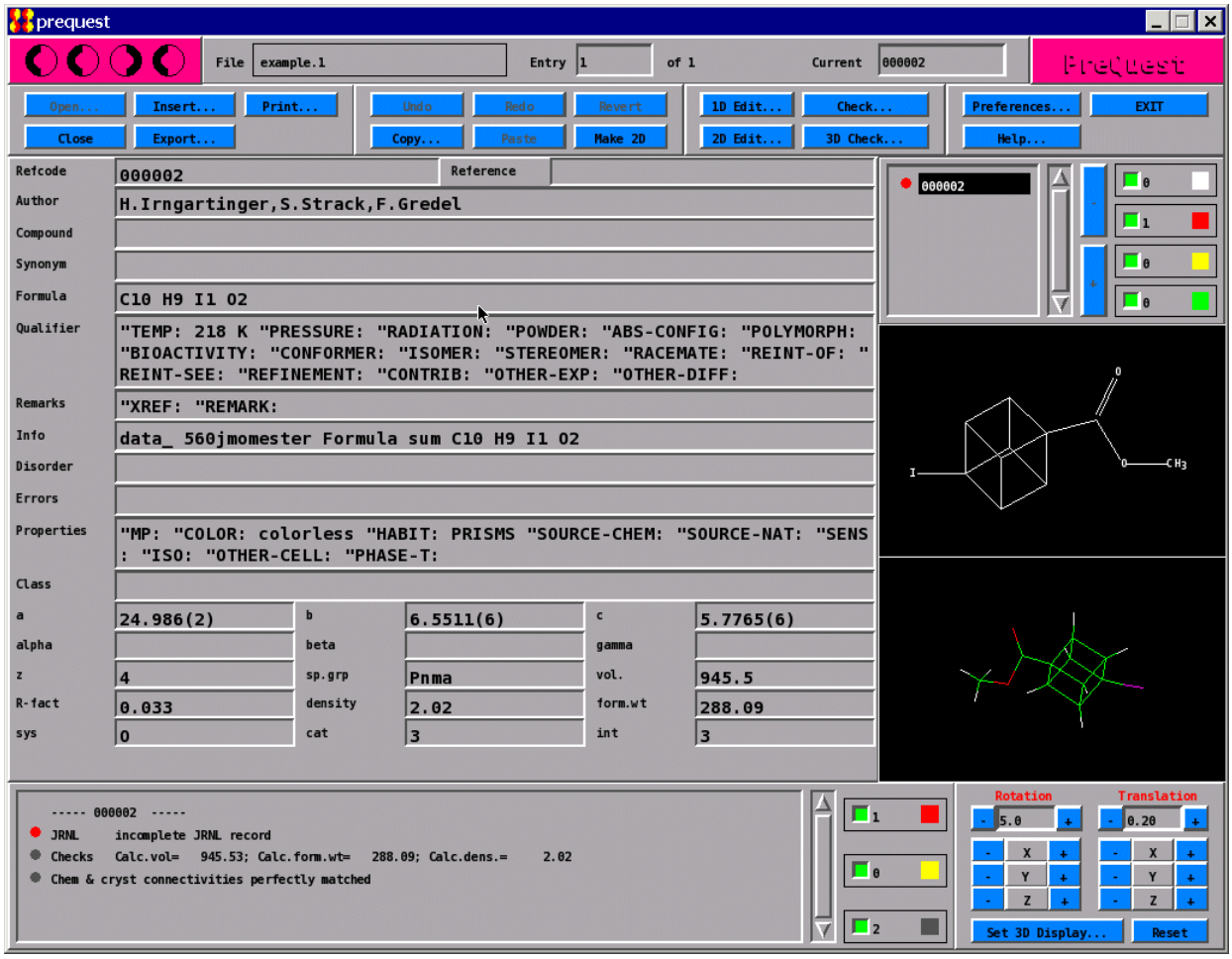
- In order to show how PreQuest can be used to create a small private database we have provided some example files with the distribution program (see *$CSDHOME/examples/prequest*). The most common formats presented by users are CIF (*example.[1-4]*) and SHELX (*example.[5-8]*). This section will illustrate what you will need to do to process these examples into a useful CSD searchable database.

2.1 A Typical Example

- start-up PreQuest, specifying X-Windows and the input file *example.1*:

```
prequest -term x -if example.1
```

- the input file is recognised as being *CIF* (see Section 5.1, page 34), and the *Main Menu* (see Section 3, page 9) appears with the example loaded:



- the program has automatically created a 2D chemical diagram from the given 3D coordinates (see Section 3.14, page 16). This diagram is essential for the search routines of the ConQuest program to function correctly, as it provides the chemical connectivity, bond types and atomic

charges. This diagram does not need to be elegant, but it can be visually improved by use of the 2D Edit function if required (see Section 3.16, page 18).

- a small amount of editing is required for this entry before it can be stored in a database. The **Message Box** contains colour-coded error messages where:
  - red - *serious*: errors should be checked
  - yellow - *warning*: should be considered
  - grey - *information only*: does not require attention
- Note: the level of checking can be controlled (see Section 2.5, page 7).
- The red error message for this entry states that there is an *incomplete JRNL record*. No journal reference was given in the input file. Structures that do not have valid journal references can be recorded as *Private Communications*. In order to classify this structure as a Private Communication, use the mouse to click anywhere in the **Reference** text box, then at the cursor type `1078,,,1996` where 1078 refers to the journal code number for Private Communications, and 1996 is the year. See BCCAB Format (see page 41) for a full explanation of the format of the reference line #*JRNL*). Hit `Return` and the red error message will be replaced by the grey information message stating that the journal reference has successfully been set to a Private Communication.
- The compound name should also be given, despite the absence of this not being an error. Use the mouse to position the cursor in the **Compound** text box, then type the compound name for this structure: `Methyl 4-iodo-1-cubane carboxylate`.
- In the *Refcode List* the entry will now have a green *Checking Status*. This means that there are no significant errors and this structure can be written to a searchable database file. To save the entry to a private database select **Export**, chose the *ASER* **format**, then type a filename (your intended database name e.g. *mybase*). Once successfully exported the database can be searched by ConQuest.

2.2 More about Editing

- PreQuest performs a large number of checks on the data fields of every entry. For convenience these can be divided into:
  - **1D Edit** (see page 4): Text and Numerical Items
  - **2D Edit** (see page 4): Chemical Diagram
  - **3D Check** (see page 4): Crystallographic Connectivity
- The individual data fields in each of these distinct classes can be easily edited from within the PreQuest program. This can be done using the *Editing Function* buttons available in the Main Menu.

When to Edit

- As a general guide, if there is a *red* error status then some editorial action should be taken.

PreQuest will allow erroneous entries to be archived, for example when they are missing fields or contain mistakes. However, it should be remembered that unchecked data greatly reduces the value of your database, and may result in search failures when using ConQuest. It is therefore well worth the effort at this stage to create records of as high a standard as possible.

- There is an option to control the levels of error checking in the *Preferences* menu (see Section 3.19, page 28). It is normally sufficient to use the *Relaxed (minimal rules)* option.

2.2.1 1D Edit

- This allows the user to edit the actual text of the record (see Section 5.5, page 41). However, many entries can be satisfactorily changed by direct overtyping in the Data Boxes provided, e.g. Author, Compound etc.
- When editing text fields in PreQuest:
  - **CONTROL-U** erases all text
  - **ESCAPE** reverts to previous contents
  - **DELETE** deletes the character left of cursor
  - **RETURN** completes the edit
  - **TAB** moves cursor to next field

2.2.2 2D Edit

- Allows modification of the chemical diagram. Specific editing of the diagram may be necessary to ensure that chemical bond types and atomic charges are correctly assigned. Users also have the option of completely redrawing the chemical diagram. This may be required, for example, if certain chemical conventions are to be adhered to (see Section 3.16, page 18).

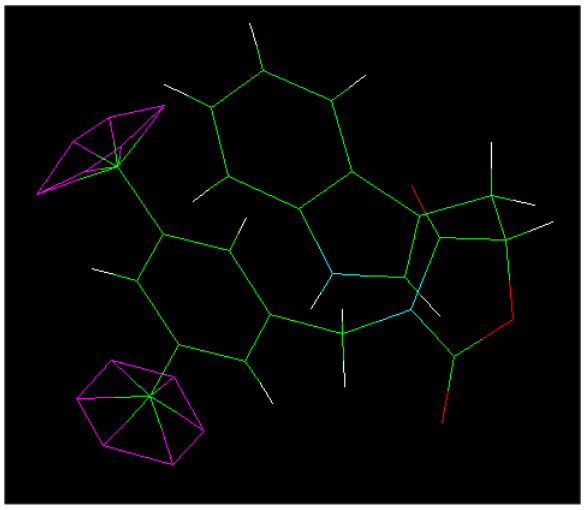
2.2.3 3D Check

- The user can change the coordinates to correct any errors in the structure, and has the ability to suppress any unwanted atomic sites in disordered structures (see Section 3.18, page 21).

2.3 Treating Disorder

- It is essential to edit the atomic coordinates for disordered sites. This is because the current version of PreQuest does not have storage facilities for disorder groups, site numbers, and occupancy factors. If a structure contains disorder (as occurs for ~20% of all CSD entries) then the minor occupancy sites must be suppressed to leave a single representative atomic position that will then be matched against the chemical diagram. Suppressed atoms are not deleted, they simply take no further part in the establishment of the crystallographic connectivity, and therefore will not be used in ConQuest 3D searches.
- How to deal with a typical example of a disordered structure is now illustrated using another example:

- Open *example.2*
- there are a number of error messages shown in the dialogue box for this structure. The summary of the number and status of these is also shown (6 red errors, 16 yellow errors). The error message about the missing journal reference can be fixed as for the first example. The remaining messages are all a result of coordinates for a disordered CF_3 group being present. This is immediately apparent from the chemical diagram which is nonsensical.



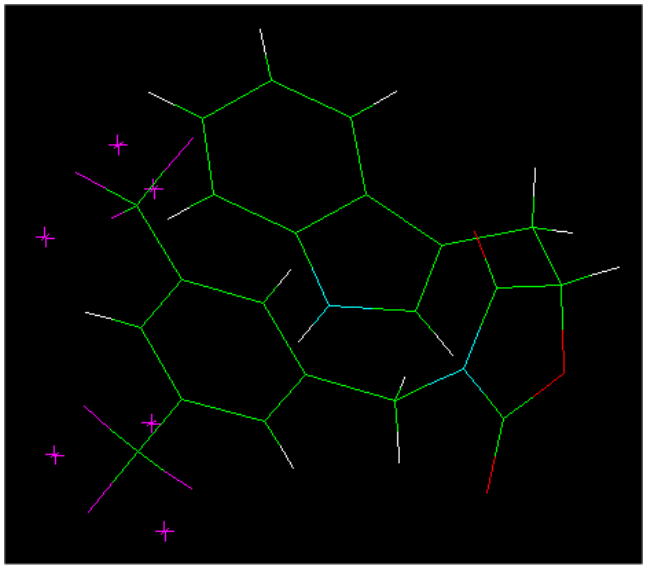
- The disorder must be treated before the entry is in a satisfactory state to be archived to a database file.
- Select **3D Check** (see Section 3.18, page 21).
- It is clear that two sites are given for each CF_3 group. Assuming that the occupancy factors are 0.5 we can choose to suppress one of those sites. Select **SUPP** (suppress) then click on the atoms F26', F27', F28', F30', F31' and F32' (you may have to use the *3D Controls* to manipulate the view of the structure). Finally re-select **SUPP** to finish the selection process. The selected atoms are suppressed - their atom labels remain on the screen but the bonds to them are removed.
- To save these edits and return to the main menu select **QuitS** (quit and save). To complete the corrections for this entry the chemical diagram needs to be re-made. Select **Make2D** (see Section 3.14, page 16). PreQuest re-makes the chemical diagram based on the new crystallographic connectivity.
- To mark an entry as disordered, so it will be excluded if the *No Disorder* filter is applied when searching via ConQuest, open the *1D Edit* window. Type the following on a new line, after the refcode line

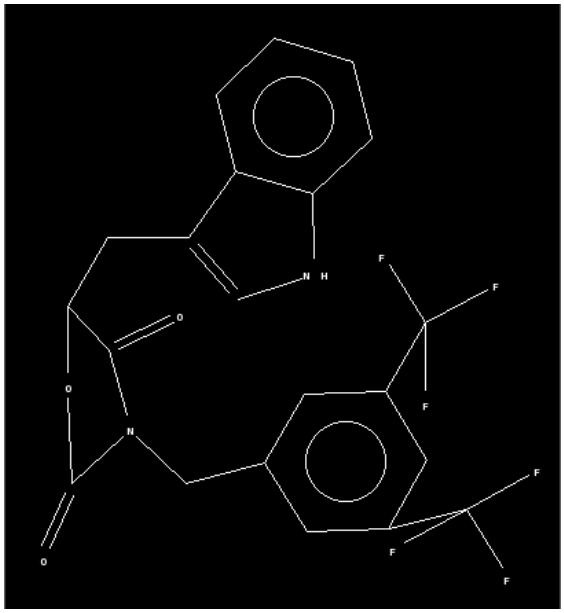
  ```
  #DISORD "COMMENT:
  ```

  then type any text to describe the disorder. This field must contain text to set the disorder flag.

- Minor occupancy sites can also be suppressed from the *1D Edit* window by adding one space then capital `S` following the final digit of the coordinate line for the atom e.g.

```
O1A 0.0178,1 0.25000 1.3046,5 S
```





- All the error messages now disappear and the entry goes *green*. This is now in a suitable state for saving to a database file.

2.4 Control of Refcodes

- The assignment of reference codes (refcodes) for your private database is controlled by an auxiliary file called *prequest.refcodes*. This file should be present in your root directory. If it is not present, PreQuest will assign a 6-digit number in sequence starting at 000001.
- *prequest.refcodes* consists of 8 lines, each of which defines the sequence of characters to be assigned by each new data entry read. This example shows the 8 lines as:

```
S
0123456789
0123456789
0123456789
0123456789
0123456789
<blank>
<blank>
```

- This has the effect of always keeping character 1 as *S* and characters 7 and 8 as *S*. The sequence of the codes will then be S00001, S00002, S00003 etc.

- Between sessions, PreQuest records the last refcode it generated in the 9th line of this file. A filename other than *prequest.refcodes* can be specified by using the environment variable `CCDCNEWREFCODES`.
- It should also be noted that the CIF field, `_database_code_CSD` can be used from within a CIF (see Section 5.1, page 34) to specify a refcode for the structure as it is read into PreQuest.

2.5 Levels of Checking

- At the CCDC production of the main database requires that all entries pass a series of detailed validation checks. This ensures that the data is of the highest standard and integrity. PreQuest incorporates two levels of checking that reflect this: *Strict (full CCDC rules)* and *Relaxed (minimal rules)*. Most users will find that the *Relaxed* setting is adequate for private database creation, and this is the default setting in the Preferences menu (see Section 3.19, page 28).
- However, some words of warning are needed. It is, for example, not necessary to give an author or compound name, but this means that these records will never be retrievable by ConQuest using these search parameters. The absolute minimum for an entry is to have a journal reference. If you switch on the *Strict* setting you will find that the author, compound name, formula and cell are also required at the CCDC.
- The *Check* menu (see Section 3.17, page 20) presents options to switch on/off checking at 1D, 2D and 3D levels. We advise keeping all of these checks **ON**. Switching off these checking tests can result in important errors being missed which result in making the entry unsearchable by ConQuest.

2.6 Recommended Minimum Data Fields

- The recommended minimum data for private databases are:
  - Journal
  - Author
  - Compound Name
  - Cell, Space Group, z-value
  - R-factor
  - Atomic Coordinates
- Note that for space groups which require exact cell angle values (e.g. for $P2_12_12_1$ the cell angles alpha = beta = gamma = 90 degrees) the alpha, beta and gamma data fields for the entry may show no values. This is because they are exact and are therefore considered *redundant* or *assumed* for that system (see Section 5.5, page 41).
- We also recommend that any textual information useful for retrieving items within your local context be added. For example:

#SYNONYM	Compound 2317P	Lab number 567894
#QUAL “BIOACTIVITY:	antibiotic activity	
#RMARKS “REMARK:	Refinement incomplete - see file *wxyz.res*	
#PROPS “SENS:	air-sensitive	

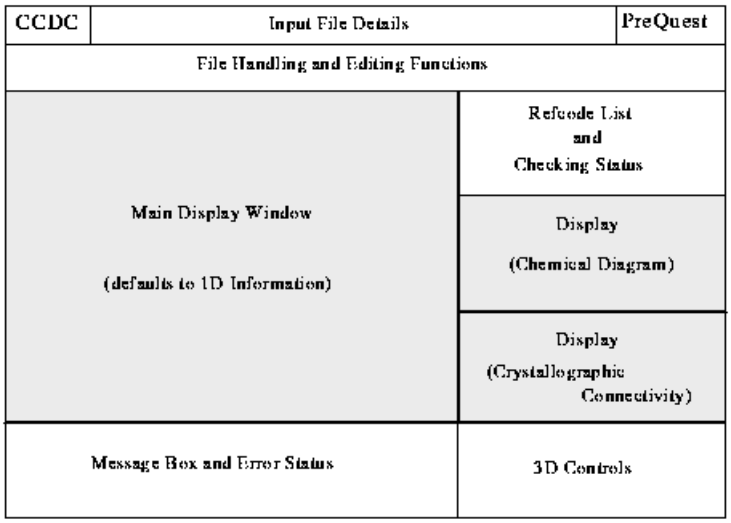
- Most of the data fields that are presented to the user have obvious meanings. They relate to the data fields of the **BCCAB** format specification (see Section 5.5, page 41). Of the slightly less obvious 1D data fields:
  - *sys* is the crystallographic system (see #SYSCAT : Crystal System and Entry Category, page 53) e.g. sys T = tetragonal
  - *cat* is the determination category (see #SYSCAT : Crystal System and Entry Category, page 53) e.g. cat 3 = full structure determination
  - *int* refers to the intensity measurement (see #UNIS : Processing Information, page 54) e.g. int 3 = diffractometer data
  - *class* refers to the chemical class assignments made for the compound (see #CLASS : Chemical Class, page 47)

3 MAIN MENU

- The main menu for PreQuest is visible upon entry to the program. This is the central menu for the PreQuest program and contains all of the control functions used for file handling, setting preferences, and checking the vailidity of the entries.

Overview

- The main menu is segregated into distinct functional parts, as illustrated by the schematic diagram below. The functions and operations available in these sections are fully described in this chapter.



Schematic diagram of the main menu for PreQuest
Default display window settings are shown in parentheses

- The most important of the file handling and editing functions are collected together. These can be generally categorised as follows:

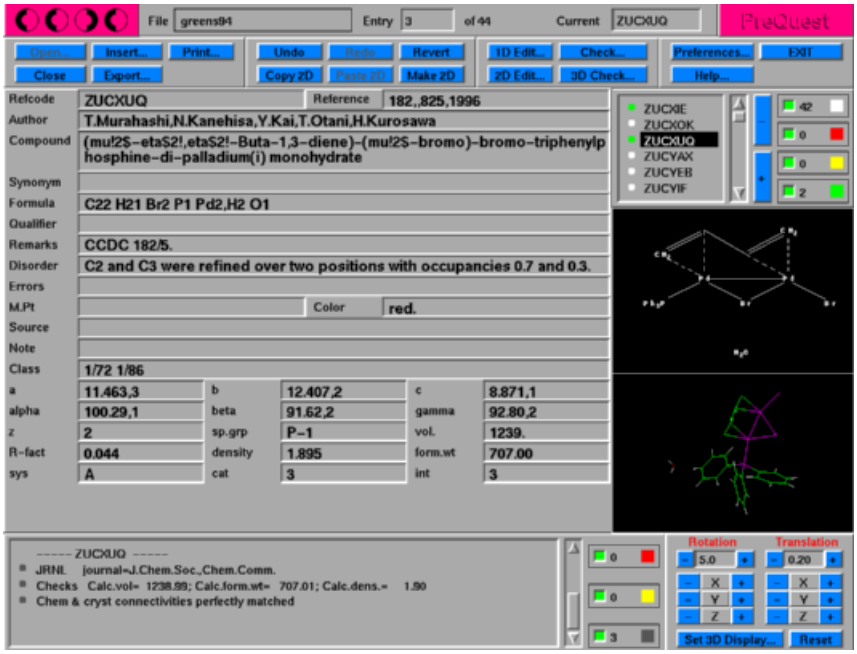
- **File Handling:**
  - **Open** (see page 12)
  - **Close** (see page 12)
  - **Insert** (see page 13)
  - **Export** (see page 13)
  - **Print** (see page 14)

- **Editing:**
  - **Undo** (see page 15)
  - **Redo** (see page 15)

- **Revert** (see page 15)
- **Copy** (see page 15)
- **Paste** (see page 16)
- **Make 2D** (see page 16)
- **1D Edit** (see page 17)
- **2D Edit** (see page 18)
- **3D Check** (see page 21)

- **Preferences:**
  - **Check** (see page 20)
  - **Preferences** (see page 28)
  - **Set 3D Display** (see page 30)
  - **Reset** (see page 31)

- **Others:**
  - **Help** (see page 29)
  - **Exit** (see page 31)

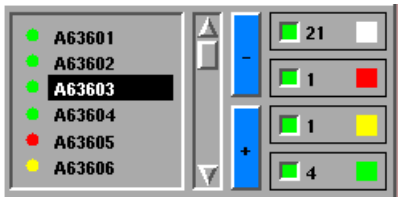


- At the bottom of the main menu window is the message area. PreQuest will display three types of colour coded message in this area:
  - *grey*: for information purposes only
  - *yellow*: warning message

- *red*: error detected by CSD validation checks
- The panel to the right shows how many messages there are of each kind. The messages may be toggled on/off by clicking in the check boxes. The types of checking performed by PreQuest can be controlled using the options available with **Check** (see Section 3.17, page 20).

3.1 Moving between Entries

- PreQuest can be used to process one data entry at a time, or to process several entries collected together into one input file. The input file details appear at the top of the main menu window. Listed here is the current input file, the name of the refcode being viewed, and that refcode's relative "position" in the file (e.g *Entry 3 of 44*).
- A scrolling refcode list is also available. This shows all of the entries in the current input file. The current entry will appear highlighted in this list.
- Refcodes can be selected for viewing simply by selecting them with the mouse. Alternatively the + and - buttons can be used to move stepwise through the input file list.



- The refcodes are colour coded by PreQuest according to their present checking status:
  - **red**: serious problem with the entry detected
  - **yellow**: checking has produced some warning messages
  - **green**: entry has been checked and has no problems
  - **white**: entry has not yet been checked
- By default an entry will be checked when it is selected for viewing. The level of checking will depend on the options currently selected in the *Check* sub-menu (see Section 3.17, page 20).
- The checking status of the whole input file is given to the right of the refcode list. These boxes give a summary of the input file information in terms of entries that have been checked and their subsequent status.
- These are toggle boxes and clicking on the Yellow box will add/remove the entries with yellow checking status from the refcode list. These refcodes are not physically removed from the file, merely removed from the current refcode list view.

3.2 Viewing Entries

- There are three areas of the main menu window that display all of the aspects for the current entry. The main area displays the 1D information (text and numerical information) by default.

This can be changed by clicking on either of the two smaller display windows, which initially default to show the 2D chemical diagram, and the 3D structural diagram.

- The 3D representation of the molecule can be manipulated using the *3D Control* buttons. Manipulation does not require the 3D representation to be in the main window.
- *Molecule Rotation*: there are 6 rotation buttons for the independent rotation about each of the three axes. The rotation increment is displayed above, in a text field. The value can be changed by the -/+ keys, which will increment the current value by 1.0 degree. Alternatively, the user can type directly into the field. Values between 0.0 degrees and 180.0 degrees are permitted.
- *Molecule Translation*: the translation buttons are analogous to the rotation buttons. The translation increment is 0.1Å, with limits of 0.0 and 2.0Å.
- Aspects of the 3D structural diagram can be changed using the options available in the *3D Display* sub-menu. These include changing the labelling, the colour of the bonds and excluding specified atoms from the view.

3.3 On-screen editing

- PreQuest features an on-screen editing function for the information visible in the 1D window. Simply select the point of insertion using the mouse and perform your edits. Selecting another field, or hitting the carriage-return key will write the edits that have just been made to the file. These edits can be undone iteratively (see Section 3.9, page 15).

3.4 Open

- Use **Open** to select an input data file. The appearance of the pop-up window will vary depending on the platform being used, however for all platforms the window contains a scrollable list of the files in your current selected directory. Click on a file and then click **OK** to import the file. Alternatively, type the file name directly into the *Directory* and *File* boxes.
- PreQuest automatically identifies the import file format. The current list of supported formats includes (see Section 5, page 34):
  - *ASER*: CSD database format (your private database)
  - *BCCAB*: CCDC internal format
  - *CIF*: Crystallographic Information File
  - *SHELX*: Output from *SHELX* crystallographic solution programs
  - *MOL*: MDL MACCS Molfile
  - *MOL2*: Tripos SYBYL file

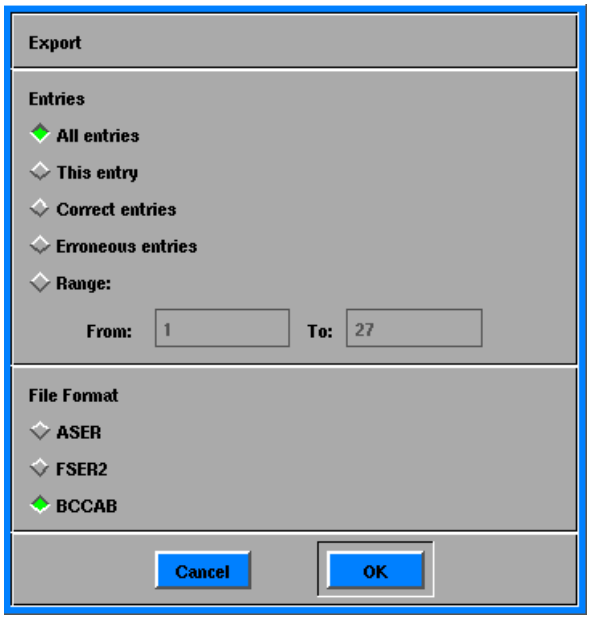
3.5 Close

- Closes the file that is currently open, but keeps PreQuest running so that it is possible to then open another file using the same session.

3.6 Insert

- Inserts a new file at the end of the currently open file, effectively concatenating the entries.

3.7 Export

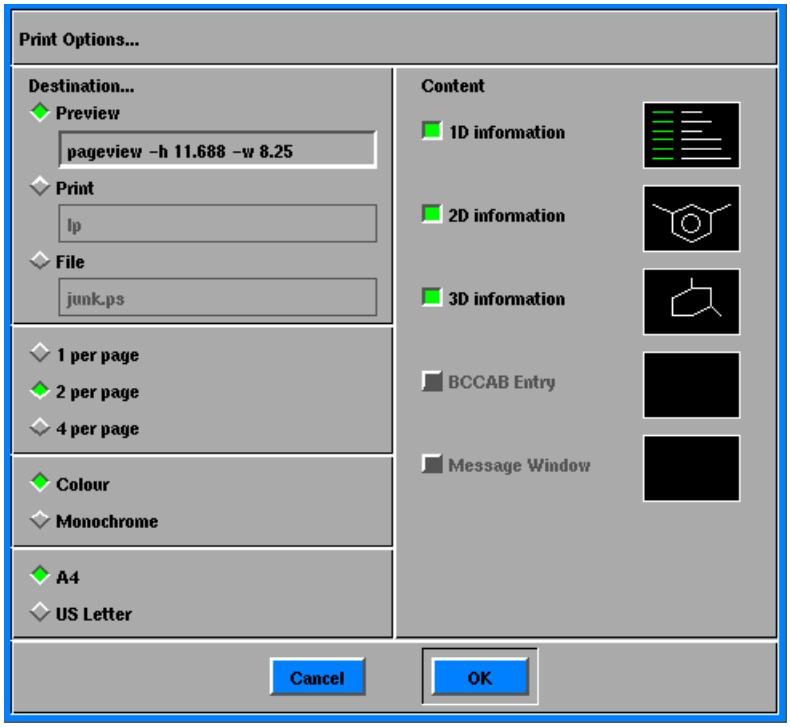


- Writes one or more of the current entries to an external file. The radio buttons allow you to select one of the following export options:
  - *All entries*: export all entries regardless of status
  - *This entry*: write the current entry only
  - *Correct entries*: export error-free (green) entries only
  - *Erroneous entries*: export erroneous (red) entries only
  - *Range*: writes the entries specified by number in the *From* and *To:* boxes. To alter the export range click in the boxes and retype the numbers.
- There are currently three types of output file format available. These can be changed by clicking on the desired button. Full format descriptions are given in Section 4.
  - *ASER*: CSD database format (your private database)
  - *BCCAB*: CCDC internal format
  - *FSER*: A sequential ASCII version of the ASER file.
- Selecting *OK* invokes a pop-up window with your current directory details listed. Select a file to save as, or type in a new name.

Which format to use?

- Normally you will output your entries into a private database using the ASER file format. If you are creating a new database simply input the desired name, (e.g. *project1*) and all of the necessary database files will be created automatically.
- If you have problem entries that are not yet ready for writing to the database, use the BCCAB format. This exports a plain text file which you can return to at a later date and proceed with PreQuest (e.g. *problem1.bcc*)

3.8 Print



- Aspects of the current entry can be printed using the Print function. A pop-up menu provides the user with the opportunity to choose which options are desired.
  - *Preview*: Postscript file sent to viewing program such as *pageview*.
  - *Print*: Type the print command required to send the Postscript output to a selected printer.
  - *File*: Postscript output written to a file with the default name shown. To change the file name simply type in the text box.
  - *1 per page*: Prints one Postscript frame per page
  - *2 per page*: Prints two Postscript frames per page
  - *4 per page*: Prints four Postscript frames per page
  - *Colour*: Select colour output
  - *Monochrome*: Select monochrome output
  - *A4*: Select A4 paper size

    - *US Letter*: Select US Letter paper size
- Check boxes toggle output on/off for the following types of output:
    - *1D information*: 1D text output
    - *2D information*: 2D chemical diagram output
    - *3D information*: 3D output of current view
    - *BCCAB entry*: BCCAB format printout
    - *Message window*: Text in the message window is printed
- Select **OK** when ready to print.

3.9 Undo

- PreQuest keeps a record of up to 200 editing steps performed on the current entry. Selecting Undo iteratively undoes each of these steps, starting with the most recent edit. A bleep will sound if no previous edits can be found.
- Individual editing steps include edits performed on the contents of the 1D window. Changing the contents of the author field, for instance, counts as an edit that can be undone.

3.10 Redo

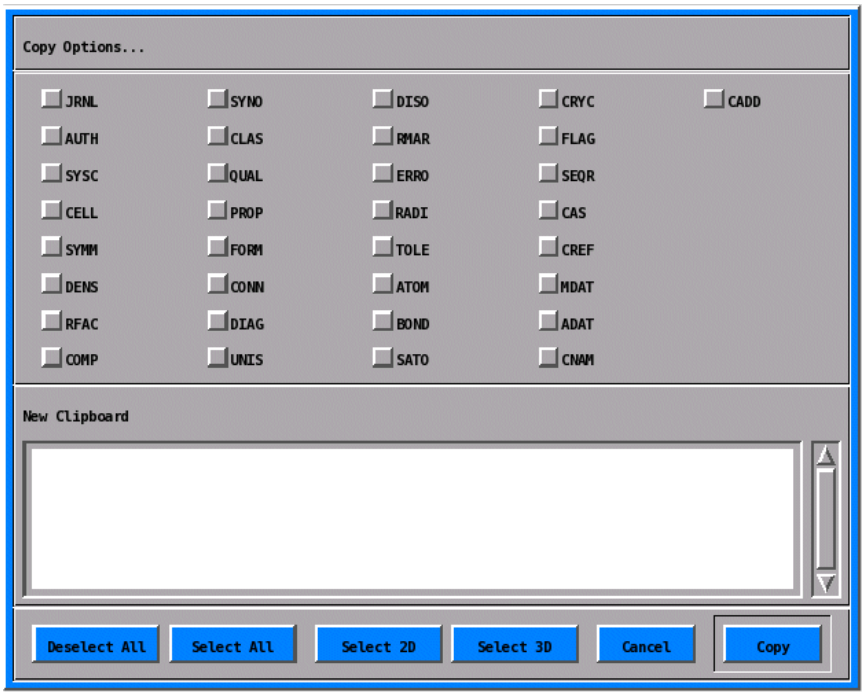
- Allows you to step forward one edit-step if you have just used Undo and made no other changes to the entry. If no forward step is available a bleep will sound.

3.11 Revert

- This function allows you to abandon all work done on the current entry and revert to the original state. Use with caution as *Revert* cannot be undone and all edits made for the entry will be lost.
- You will be prompted for affirmation of your intent to discard all edits should *Revert* be selected. Choose *Revert* again if you wish to proceed or *Cancel* should you wish to return to your edited entry.

3.12 Copy

- A useful feature for copying 2D chemical diagrams or other data for a series of similar compounds.
- Click on **Copy** then activate check boxes depending on which data you wish to copy. Note that to copy a diagram, *DIAG* and *CONN* boxes must both be checked.



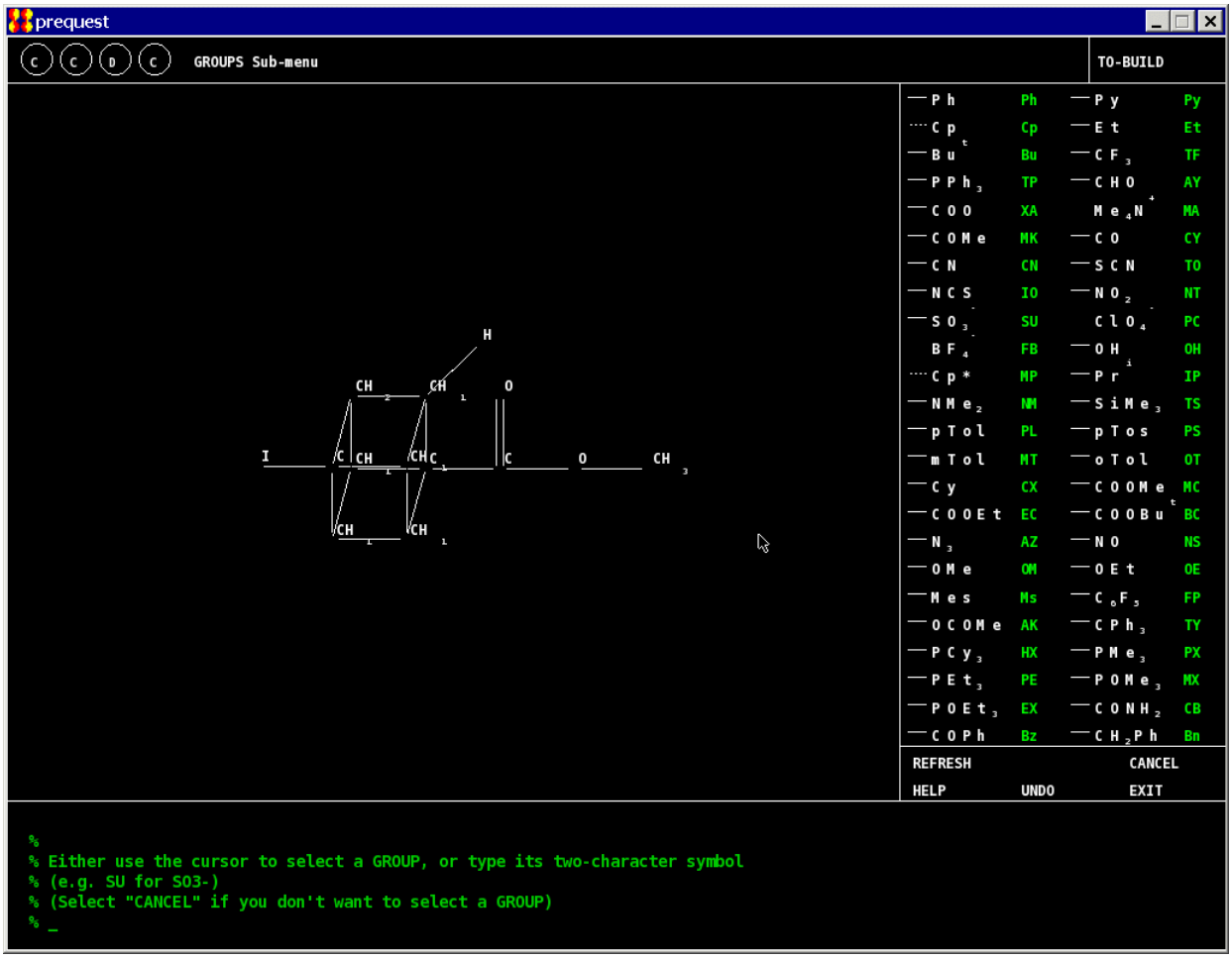
- Various selection options are provided at the bottom of the window.
- Click the **Copy** button to finish the selection and close the window.
- Data are copied to the PreQuest clipboard and can be pasted back in when required.

3.13 Paste

- Pastes the current clipboard contents into the entry. Note that the **Paste** button will be greyed out if there are no data to paste.

3.14 Make 2D

- This is a very powerful feature for automatically creating the 2D chemical diagram and the chemical formula from the current display of the 3D structure.
- PreQuest automatically assigns chemical bond types according to CSD standard practices (e.g. all nitro groups will be coded the same way). Each chemically discrete bonded molecule is taken separately and a search for the minimum overlap view is performed. The molecules are then arranged in a vertical stacked manner. The diagrams may be simplified by use of the CSD standard group symbols e.g. Ph for phenyl (see Appendix 1: Group Symbols, page 55), (see 2D Edit, page 18).



- The process is highly successful for organic compounds and will even work in the complete or partial absence of hydrogen atoms. There may be occasional ambiguities, and some metal complexes and delocalised-charge systems are known to give problems, so these should always be visually checked in the 2D window.
- If the diagram is not acceptable it can be edited using **2D Edit**.
- ***Important Note***: It is hoped that most users will find the 2D diagrams acceptable for their private database. However, it is important to realise that this chemical connectivity is vital to the sub-structural search procedures in ConQuest. The bond types and charges will be exactly as you provide them. Any 3D search in ConQuest requires that a hit is ***first made on the 2D diagram***.

3.15 1D Edit

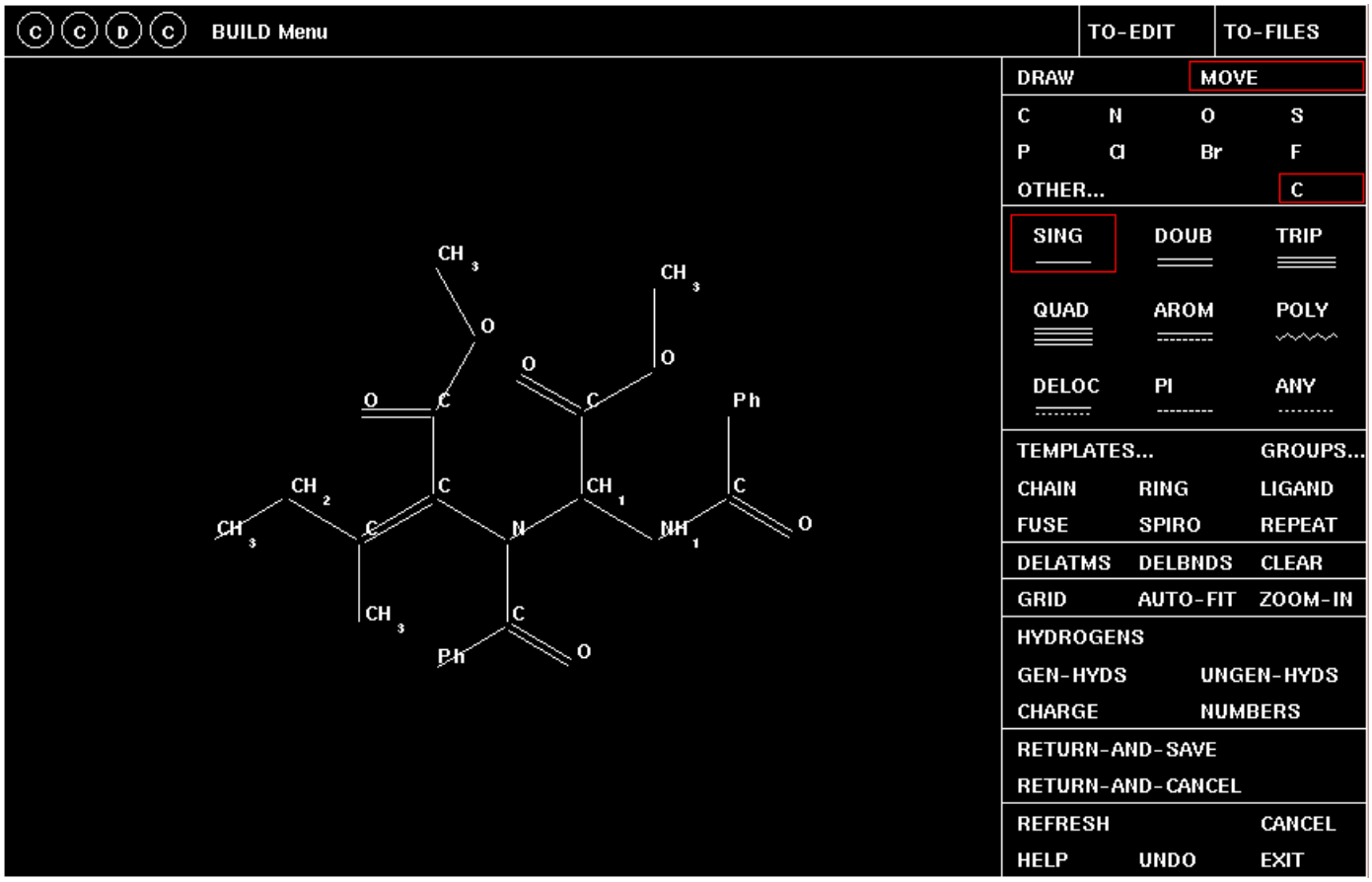
- This command sends the current entry, in BCCAB format, to the text editor selected in the *Preferences* menu (see Section 3.19, page 28).
- The text editor will launch a window which is then available for manually editing the entry. Those PreQuest commands which could modify the current entry are forbidden until the editing session is exited. When you have finished and quit from the text editor the new BCCAB text is

read back into the program, and all active checks are re-run.

- The BCCAB format is designed for easy text editing and is defined in Chapter 4 (see Section 5, page 34).
- ***Note***: all the text fields visible in the 1D window can be directly edited by selecting the insertion/edit point with the mouse and proceeding from there. The 1D Edit function is, therefore, only really necessary where the fields that need to be changed are hidden. Further information can be found earlier in the documentation (see Section 2.2.1, page 4).

3.16 2D Edit

- This button launches the 2D chemical diagram editor which allows you to modify the 2D diagram, or even draw the diagram from scratch.
- The 2D Editor menus are very similar to those described in detail in the Quest manual (Vol 2, *chapters 1*, *chapter 5* and *chapter 6*) so only a brief guide is given here. Some features of the editor will be rarely used as it is possible to draw most diagrams using only a small subset of commands.



- **DRAW**: When active this draws a bond from the currently highlighted atom to the cursor position.

- **MOVE**: When active this draws no bond but adds a new atom at the cursor position.
- C N O S etc.: To change current atom type click on symbol. Other allows element selection from a periodic table menu. Any element symbol can be typed into the element box.
- **SING, DOUB** etc.: Click on these symbols to set the current bond type.
- **TEMPLATES**: A set of template structures are available for ease of drawing. A list detailing some of the templates provided can be found in the Quest documentation (see Chapter 3: *http://www.ccdc.cam.ac.uk/support/documentation/quest/volume3/z3toc.html#9*).
- **GROUPS**: Allows selection from a set of groups symbols permitted in the CSD e.g. Ph for phenyl (see Appendix 1: Group Symbols, page 55).
- **CHAIN**: Creates a chain of the currently selected elements.
- **RING**: Creates an N membered ring of currently selected atom and bond types.
- **LIGAND**: Specialised use for attaching groups to metal complexes.
- **FUSE**: Fuses two rings together at a specified position.
- **SPIRO**: Performs spiro fusion of two molecules.
- **REPEAT**: Do not use.
- **DELATMS**: Deletes atoms as selected. Toggle off to finish deletions.
- **DELBNDS**: Deletes bonds as selected.Toggle off to finish deletions.
- **CLEAR**: Clears the drawing area completely. Cannot be undone.
- **GRID**: Draws a grid to assist tidy drawing.
- **AUTO-FIT**: Move and scale the drawing so that it just fits into the window.
- **ZOOM-IN**: Zoom into drawing to enlarge a specific area.
- **HYDROGENS**: Sets the number of terminal hydrogens.
- **GEN-HYDS**: Automatic generation of terminal hydrogens on carbons using bond type information.
- **UNGEN-HYDS**: Removes all terminal hydrogens from carbons.
- **CHARGE**: Type number for charge (e.g. -1) then select atoms to place charge on.
- **NUMBERS**: Displays sequential numbers beside atoms.
- **RETURN-AND-SAVE**: Returns to main menu saving the current 2D diagram.
- **RETURN-AND-CANCEL**: Returns to main menu discarding any edits made in the 2D Editor. The original 2D diagram is restored.
- **REFRESH**: Refreshes the screen.
- **CANCEL**: Cancels a command selected by mistake.
- **HELP**: Select a button for a brief description.

3.16.1 EDIT menu

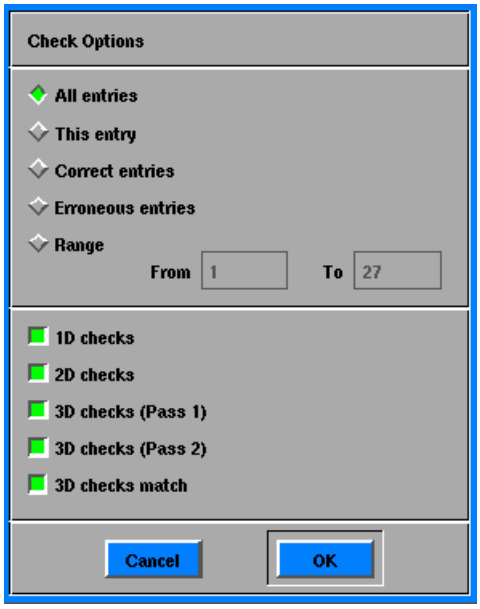
- This menu contains many graphical editing options which are intended for construction of high-

quality symmetrical diagrams. The full description of these functions is beyond the scope of this User Guide and the interested user is referred to the Quest documentation (*Vol 2 Chapter 5*).

3.16.2 FILES menu

- The *Files* menu provides features for storing and fetching fragments for 2D diagrams. There are also facilities for input, generation or changing of the chemical formula.
- Before leaving the *2D Edit* window, the formula for the structure must be generated. To do this click on **GENERATE-HYDS** (this will add hydrogens to give standard C valency so is useful where all hydrogens were not generated by *MAKE2D* or if bond types have been manually changed). Then click **GENERATE-FORMULA**, which adds a formula calculated from the 2D diagram. Finally, hit **RETURN-AND-SAVE** to exit the *2D Edit* window.

3.17 Check

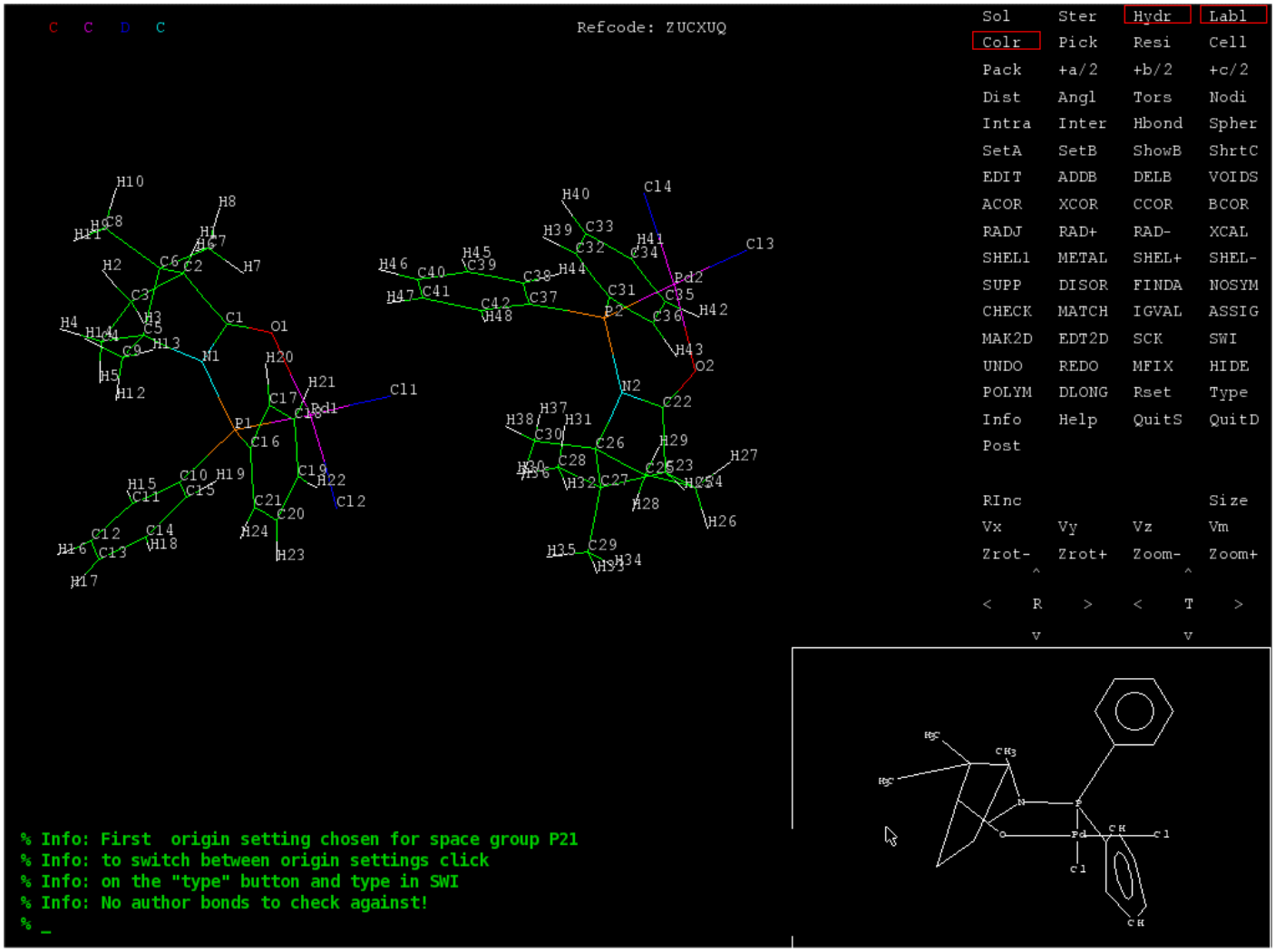


- Use the radio buttons to control the range of entries to be checked:
  - **All entries**: Apply checks to all entries in file
  - **This entry**: Apply checks only to this entry
  - **Correct entries**: Apply checks only to correct (green) entries
  - **Erroneous entries**: Apply checks only to yellow/red entries
  - **Range**: Apply checks only to entries numbered
  - **From**: First entry in range (overtype in box)
  - **To**: Last entry in range (overtype in box)
- Use the following check boxes to specify the level of checking:

- **1D checks:** Apply *CCDC* checks on 1D information fields e.g. compound names, author, cell dimensions, space group, etc.
- **2D checks**: Apply CCDC checks on the 2D fields e.g. consistency of 2D chemical diagram with the chemical formula, valency of atoms, etc.
- **3D checks**: Apply CCDC checks on 3D fields.
- **(pass1)**: The first pass checks for internal consistency of the 3D data, e.g. duplicate atom names, atoms in #BOND but not in #ATOM.
- **(pass 2)**: The second pass compares any author given bond lengths (#BOND), with values calculated by PreQuest, flagging any discrepancies. Abnormal valencies are detected and unreasonable H-C-H angles are noted.
- **3D match**: This compares each molecule of the 3D structure against the 2D chemical diagram. If any molecule fails to match the status is set to "partial match".
- **OK**: Apply the checks as set above.

3.18 3D Check

- This menu allows the checking and editing of the 3D coordinates and connectivity. It is based on the *RPluto* program but contains extra editing features designed to cope with the extensive checking required at the CCDC for creation of the main CSD file. It is unlikely that most users will need to use many of these features, however a brief description of their use now follows.
- On-line Help is available for each button.


Refcode: ZUCXUQ
Sol Ster Hydr Labl
Colr Pick Resi Cell
Pack +a/2 +b/2 +c/2
Dist Angl Tors Nodi
Intra Inter Hbond Spher
SetA SetB ShowB ShrtC
EDIT ADDB DELB VOIDS
ACOR XCOR CCOR BCOR
RADJ RAD+ RAD- XCAL
SHEL1 METAL SHEL+ SHEL-
SUPP DISOR FINDA NOSYM
CHECK MATCH IGVAL ASSIG
MAK2D EDT2D SCK SWI
UNDO REDO MFIX HIDE
POLYM DLONG Rset Type
Info Help QuitS QuitD
Post
RInc Size
Vx Vy Vz Vm
Zrot- Zrot+ Zoom- Zoom+
% Info: First origin setting chosen for space group P21
% Info: to switch between origin settings click
% Info: on the "type" button and type in SWI
% Info: No author bonds to check against!
% _

- The range of **Display Functions** available include:
  - **Sol**: Toggles between ball-and-spoke (solid) and stick drawing.
  - **Ster**: Stereo pairs or single diagram.
  - **Hydr**: Toggles display of hydrogens.
  - **Labl**: Toggles labelling of atoms: all or none.
  - **Colr**: Colour by elements or monochrome.
  - **Pick**: Pick an area of the diagram by polygon enclosure for zooming in on details.
  - **Resi**: Select a single residue for display in cases of complex structures with multiple residues.
  - **Cell**: Draws the unit cell.
  - **Pack**: Draws all molecules having a centre of gravity within the unit cell axial range 0-10
  - **+a/2**: Extends the range for packing by +a/2
  - **+b/2**: Extends the range for packing by +b/2
  - **+c/2**: Extends the range for packing by +c/2

- **Crystallographic Properties** relate to the set of operations which can be performed on the structure based on the inherent geometrical and spacial information available.
  - **Dist**: Select any 2 atoms and the distance between them will be displayed on the screen. Double click Dist to display all molecular bond lengths.
  - **Angle**: Select any 3 atoms and the angle between them will be displayed on screen.
  - **Tors**: Select any four atoms to display the torsion angle on the diagram.
  - **Nodi**: Clears the display of all distance, angle, and torsion values.
  - **Intra**: Recalculates the intramolecular geometry using the standard CSD covalent radii table and tolerance values, unless these have been re-defined in the BCCAB #RADIUS field (see Quest documentation, Appendix 10: *http://www.ccdc.cam.ac.uk/support/documentation/quest/volume3/z310.html*).
  - **Inter**: Calculates and displays all intermolecular contacts found to be less than the sum of the standard CSD van der Waals radii (see Quest documentation, Appendix 10: *http://www.ccdc.cam.ac.uk/support/documentation/quest/volume3/z310.html*).
  - **Hbond**: Intermolecular contacts are shown when the non-bonded distance is less than the sum of the van der Waals radii for the limited set of atoms C, N, O, F, Cl, Br, I.
  - **Sphere**: All intermolecular contacts are displayed within a sphere of given radius. Click on the required atom and then type a radius.
- The **Editing and Checking Functions** available allow the user to perform a number of corrective operations. The primary objective is to get the structure to pass all of the PreQuest checking routines, and also for the crystallographic connectivity to match that given in the 2D chemical diagram.
  - **EDIT**: Invokes the text editor in the same way as in the main menu. A text editing window appears displaying the full BCCAB format record. After making any changes and quitting the window the program re-runs all the 3D checks and displays the structure.
  - **ADDB**: Adds a bond between any two atoms by clicking on those two atoms. These extra bonds, not generated automatically by the covalent radius method, are treated as normal bonds and are searchable by ConQuest.
  - **DELB**: Deletes a bond between any two atoms. This is most commonly used to remove extra metal-ligand bonds from metal complexes which cannot be corrected by reducing the atomic radii. If any other function involving recalculation of bonds is performed subsequently (eg using **RADJ** or **POLYM** functions) added or deleted bonds will not be retained and the **ADDB** or **DELB** routine must be repeated. It is, therefore, advisable that if **ADDB** or **DELB** is to be used, it is used immediately prior to quitting the 3D window (using **Quit S-** see later).
  - **ACOR**: Allows direct setting of an atom field in BCCAB. The atom label can be changed and the individual coordinates can be altered.
  - **XCOR**: Provides a means of correcting atomic coordinates when there is a discrepancy between the calculated bonds to an atom and the author specified bonds. This situation should

be rare for private database creation, but is frequently encountered at the CCDC. PreQuest attempts to suggest a corrected coordinate assuming that the other coordinates are correct.

- **CCOR**: Provides a means of correcting mis-typed cell parameters. It uses a least squares fit of calculated versus given bond lengths.
- **BCOR**: Provides a correction method for mis-typed author given bonds.
- **RADJ**: Allows interactive adjustment of the covalent radii for a specified element. Click on the atom and then use RAD+, RAD- to adjust the radius by a specified increment.
- **RAD+**: Incrementally increases the elemental radius.
- **RAD-**: Incrementally decreases the elemental radius.
- **XCAL**: coordinate calculation by geometry, based on existing atoms in the structure. Used to complete hexagons or pentagons when an atom is obviously missing.
- **SHEL1**: Display option allowing visualization of a structure by moving outward from a given atom by "shells" of connectivity. Particularly useful for complicated metal complex problems. Used in conjunction with SHEL+ and SHEL-.
- **METAL**: Starts the SHEL1 process on all metals in the structure. Good for metal-cluster problems.
- **SHEL+**: Displays an additional "shell" of connectivity
- **SHEL-**: Displays one less "shell" of connectivity.
- **SUPP**: Suppresses an atom. A suppressed atom no longer belongs to the searchable structure although the information relating to it is kept for future possible use. Suppression is currently the only method for dealing with crystallographic disorder - minor occupancy sites are currently suppressed at the CCDC during database building. Select **SUPP**, then up to 10 atoms to be suppressed, then click SUPP again. The atoms will disappear from the display and be replaced by grey question marks. Repeat this procedure to un-suppress the atoms.
- **DISOR**: Not available.
- **FINDA**: Find atom by name and label.
- **NOSYM**: Switch off all symmetry generated atoms displaying only the asymmetric unit of the structure. This is a toggle function. PreQuest forms a bonded network by applying space group symmetry operations to the atomic coordinates. If the molecular symmetry coincides with some space group symmetry then the program adds the symmetry generated atoms. Useful for viewing molecules which are disordered by symmetry.
- **SetA**: Allows properties of an atom (i.e. charge and hydrogen count) to be set in 3D. This may be of use so that the **Make2D** diagram is correct and requires no editing. Click **SetA** then follow the prompts, **SetA** to finish.
- **SetB**: Allows properties of a bond to be set in 3D. Click **SetB** and follow the prompts, **SetB** to finish
- **VOIDS**: checks the structure for void space. Click **VOIDS** then return to use the default radius of 2.4 Angstroms, or change this if required, then return. The percentage of void space found

will be indicated.

- **CHECK**: Runs the 3D Check procedures with messages written at bottom of window (scrollable with the up-arrow, down-arrow keys). Highlighting of erroneous bonds occurs where appropriate. All calculated bonds are checked for a matching author given bond (*#BOND*). Bond "errors" are shown according to value of the discrepancy:

```
0.02 - 0.05 A = yellow
> 0.05 A = red
```

Messages for very short or very long bonds are also given. Bonds given in #BOND but not calculated show as blue dashed lines.

- **MATCH**: Attempts to match the chemical connectivity specified in the 2D chemical diagram with the connectivity generated in the 3D structural representation. Each part of the structure is flagged *MATCHED* or *NO MATCH*. Missing hydrogen atoms in the 3D structure are allowed by the matching process. Matching problems will not occur if the automatic *MAKE2D* (see Section 3.14, page 16) is used. ***Note:*** It is vital that every molecule in the 3D structure matches a molecule in the 2D diagram because the ConQuest search for 3D geometric structures will not access the 3D coordinates unless the match status is "perfect match".
- **IGVAL**: Ignore valency errors. Instructs PreQuest to ignore all of the valency errors that are encountered in checking.
- **ASSIG**: Assigns bond types to the 3D structure. This is a required procedure before Make2D can be applied to the 3D structure. The 3D structure is displayed with bond types depicted as single, double, and triple lines. Dashed lines indicate aromatic or de-localised bonds. Grey dotted lines indicate pi-bonding. Rules are applied systematically using bond lengths, angles and torsion angles. The method tolerates missing hydrogen atoms. Standard patterns are assigned following CSD database conventions e.g. a perchlorate group has three double bonds, one single bond, and a single negative charge on that single-bonded oxygen. Full details are beyond the scope of this User Guide, but in general the process is highly successful for organic structures. Some ambiguities can arise with de-localised charge systems and with some metal complexes.
- **UNDO**: Iteratively undoes each separate stage of edits. The most recent 200 edits for one structure are stored by PreQuest and each can be undone until no more steps are available. Moving to another structure removes the ability to undo edits to a previous structure.
- **REDO**: Iteratively re-applies the edits that have been undone using the UNDO function.
- **POLYM**: This feature is under development for the handling of catena-structures. There is generally an editorial decision to be made as to how much of a polymeric chain should be shown as the motif in the 2D diagram. POLYM options for use on polymeric structures:
- **Polym=1**: Show expanded polymer.

- **Polym=2**: Selects and tries to match on 2D chemical diagram.
- **Polym=3**: Select crystal asymmetric unit + 1 link.
- **Polym=6**: Label the base unit only.
- **Polym=7**: Accept this polymer unit (saves this unit, so should be used as soon as you are happy with what is displayed).
- **Polym=8**: Expand selected atoms by a set of bonds.
- **Polym=9**: Expand all atoms by one set of bonds.
- **Polym=10**: Expand by complete ligands around metals, selects all metals automatically.
- **Polym=11**: Derives a unit of polymer which has the metals and ligands in the correct proportions for the molecular formula. This is achieved by taking into account the general/ special position multiplicity of the metals and ligands. However, the particular choice of polymer unit is arbitrary. It also assigns polymeric bond types to any polymeric bonds. This unit should be suitable for use in **MAKE2D**.
- **Polym=12**: Does the same as polym=2 but with more match iterations.
- **Polym=18/19**: As 8/9. but do not restore deleted bonds.
- **Polym=20**: As previous **Polym=10**, prompts for metal atoms.
- **Polym=21**: As **Polym=11** but prompts for metal atoms (this allows option to be used when polymer does not contain atoms conventionally considered as being metals).
- **Polym=22**: Does the same as **Polym=2** and **Polym=12** but uses even more iterations (and can take a considerable time) but is useful in extreme circumstances.
- **Polym=28**: Complete selected ligand(s), by clicking on any non-metal atom in the required ligands(s). Also adds any metal atoms bonded to those ligands.
- **Polym=29**: Complete all incomplete ligands currently displayed (including those not bonded to any metals, eg solvents).

Options 28 and 29 may be useful in completing part-ligands in the base asymmetric unit (**Polym=3**) or for expanding the polymer in a controlled manner, to build up a unit which corresponds to that drawn in the 2D diagram.

It is worth noting that with monoatomic ligands (e.g. O, Cl etc.) **POLYM=11,12** options cannot always determine whether the single atom in the ligands whould be included in the formula or not. Where the formula generated has "extra" ligands, these can be removed from the formula by changing the bond types involving the ligands to all polymeric (ie in *2D Build* menu, click on polymeric bond type then click on the bond you wish to make polymeric. To assist in locating the "extra" ligands you can label the atoms in 2D with their 3D labels: type `lab3D`. This will show symmetry generated atoms with a suffix e.g. for a structure which has an "extra" water ligand, type `lab3D` gives *O1* and *O1A*. *O1A* is likely to be the "extra" atom and bonds to this atom should be changed to polymeric. To remove these atom labels type

`lab2D`.

Extra H atoms may be generated by the **POLYM** functions e.g. a polymerically bonded CH unit may be displayed as CH_3. To correct this in *2D Edit* click on **TO-BUILD**, **HYDROGENS**, type the number of H atoms required then hit return. Next, click on the atom requiring H atoms, then **HYDROGENS** to finish. Do not click on **GEN HYDROGENS** as this will recalculate as CH_3.

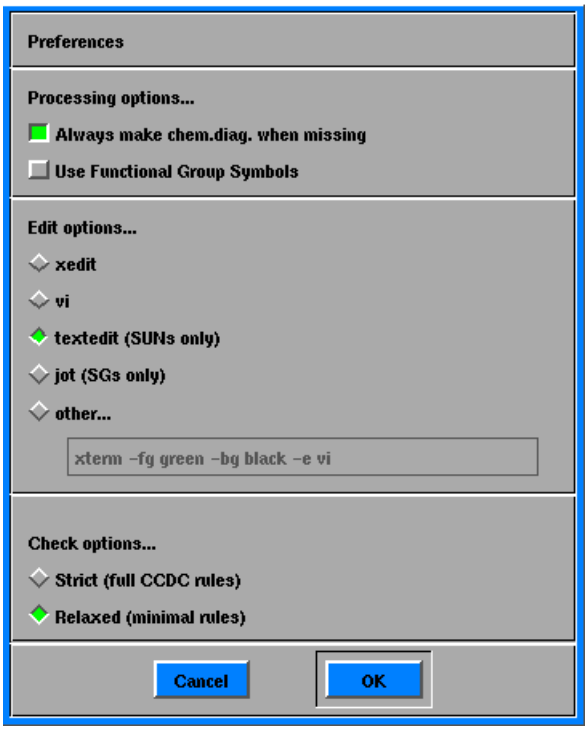
Recommended procedure for use on polymeric structures:

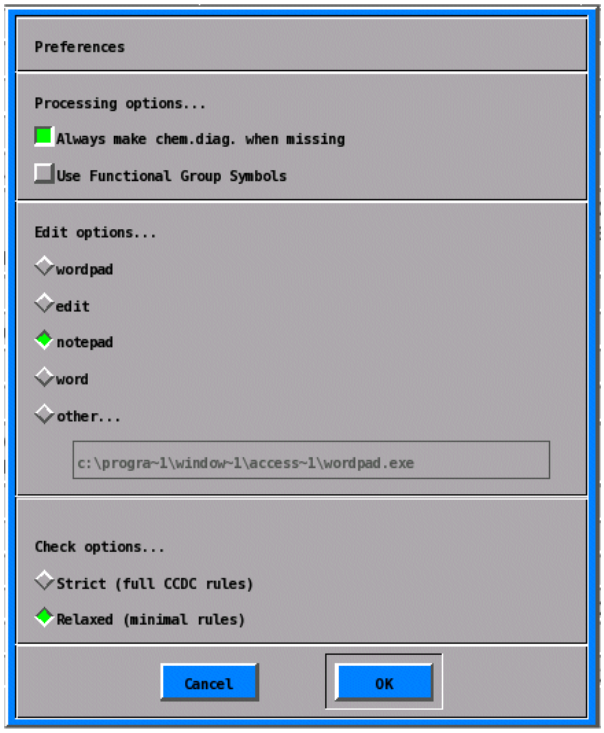
Click on ASSIG then POLYM 11 then MAKE 2D to generate 2D diagram based on this polymer unit, followed by POLYM 7 to save the polymer, POLYM 2 to match it to the 2D diagram and QUIT S to exit the 3D window.

- **DLONG**: Delete all "long bonds" given in the author's bond list (#BOND). These are defined as bond distance > (radius1 + radius2 + tolerance) for the current setting of #RADIUS and #TOLER.
- **Rset**: Resets the display to the minimum overlap view.
- **Type**: Commands can be typed in instead of using the relevant buttons.
- **Info**: Displays brief information on cell parameters, space group, Z value etc.
- **Help**: Short descriptions of the various functions are available.
- **Quit S**: Return to main menu and save the changes made.
- **Quit D**: Return to main menu and discard all of the 3D edits.
- **Post**: Output a postscript file of the current 3D display. Output to file probname.ps

- Finally, there are a number of 3D Controls which can be used to manipulate the view of the 3D crystallographic structure in the main display window
  - **Vx**: View down the cell a-axis
  - **Vy**: View down the cell b-axis
  - **Vz**: View down the cell c-axis
  - **Vm**: View in minimum overlap projection
  - **Zrot+**: Rotate clockwise about the z-axis (steps of 3 degrees)
  - **Zrot-**: Rotate anti-clockwise about the z-axis
  - **Zoom-**: Zoom outwards by scale factor of 10%
  - **Zoom+**: Zoom inwards by scale factor of 10%
  - **Size**: Scale plot to fit in main window
  - **R**: Rotate in steps of 3 degrees
  - **T**: Translate in steps of 0.5 Å

3.19 Preferences



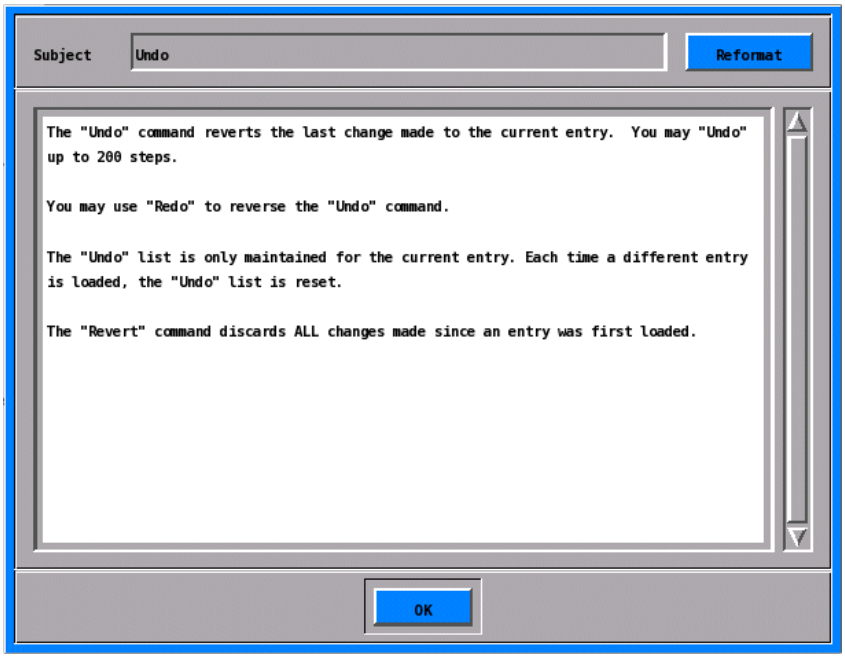


- A number of preferences are available for the user to set at their discretion. Default options have their check boxes turned on (green).
- There are two processing options available:
- **Always make chem.diag. when missing** does exactly this. If PreQuest finds no 2D chemical diagram for an entry it will automatically perform **Make2D** in order to generate one, saving the user from having to perform this operation for every entry.
- **Use Functional Group Symbols** will apply the standard set of CSD symbols when **Make2D** is used. When checked this option is applied whether Make2D is automatically invoked, or manually selected by the user.
- Choose the **text editor** preference here, the choice will vary depending on the platform you are using. Pick from one of the text editors provided, or select **other...** and enter the command to call your preferred editor. Click on the box and input the complete initialisation command.
  Note: If you want to use the **1D Edit** function on a Macintosh, PreQuest needs to be started from the command line on a terminal.
- **Check options** can also be stipulated using this menu:
- **Strict** applies all of the checks and conventions that are required to be performed at the CCDC on all entries before they are archived to the main CSD file. This includes a large number of checks on the bibliographic and crystallographic information fields which are not necessary for building a private database.
- **Relaxed** applies a less stringent set of rules to each entry. This is the default option for private

database building.

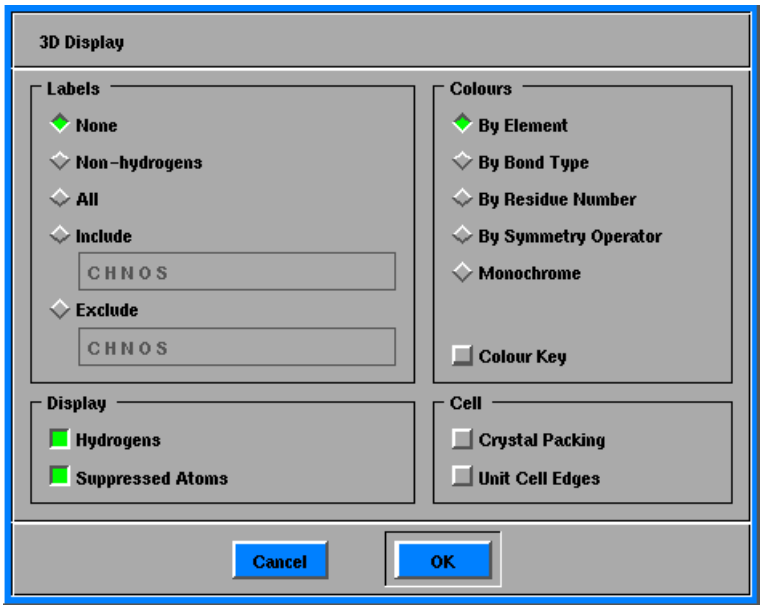
- **Save** this set of preferences for processing future entries.

3.20 Help


Subject
Undo
Reformat
The "Undo" command reverts the last change made to the current entry. You may "Undo" up to 200 steps.
You may use "Redo" to reverse the "Undo" command.
The "Undo" list is only maintained for the current entry. Each time a different entry is loaded, the "Undo" list is reset.
The "Revert" command discards ALL changes made since an entry was first loaded.
OK

- The pop-up *Help* window is launched which remains active while you click on any item on the main menu. A short description of the function or area is given in a scrollable text format. In the example above the UNDO button has been selected.
- Select **Return** to remove the Help window.

3.21 Set 3D Display



- This pop-up menu gives the options available for the setting the 3D display in the main menu:
- Label Options
  - **None:** No atom labels.
  - **Non-hydrogens:** Label all atoms except hydrogen.
  - **All:** Label all atoms.
  - **Include:** Label only those elements appearing in the text box. To change the list, click in the box and over-type the current element set e.g. Br O N.
  - **Exclude:** Label all atoms except those elements appearing in the text box. To change the list, click in the box and over-type. e.g. C H.
- Colour Options
  - **By Element:** Colour by element, as in *ConQuest.*
  - **By Bond Type:** Colour bonds by bond type.
  - **By Residue Number:** Each residue (molecule) is given a different colour.
  - **By Symmetry Operator:** Each molecule in the packing diagram is given a colour code related to the space group symmetry operator used to position it.
  - **Monochrome:** All bonds monochrome (white).
  - **Colour Key:** Toggle display of the colour key to the top right of the 3D display.
- Display Options
  - **Hydrogens:** Toggle display of hydrogens.
  - **Suppressed Atoms:** No effect.
- Cell Options

- **Crystal Packing:** Toggle packing display within the unit cell axis range of 0-1.
- **Unit Cell Edges:** Click on/off to show unit cell edges.
- **OK:** Click here to return and apply these settings.

3.22 Reset

- Resets the view of the 3D window to the default.

3.23 Exit

- This button prompts you with a reminder that you must **Export** your file prior to finishing your current session. If you do not export there is no way of recovering any changes you have made.

4 PRIVATE DATABASES

4.1 Creating a Database

- A searchable private database is created the first time that an entry is saved from *PreQuest* in the **ASER** format, with a specified name:

```
mybase.ind
mybase.msk
mybase.tcd
```

- The *mybase* database can be searched using ConQuest after it has been correctly activated, see the CSD System installation notes for further information.

4.2 Adding New Entries

- Single entries, or a file of entries can be added to a private database by selecting the relevant **Export** option and specifying the name of the database to which the entries will be added. The format must be **ASER**. For example, to add entries to the *mybase* database simply type `mybase` in the **File** text box, select the **ASER** format and hit return. PreQuest will write the records with accompanying log messages appearing in the text window.

  writing entry # n - refcode

- If a refcode matches an entry already present in the database it will not be written to the database and an error message will be issued. For this reason it is essential that each new entry has a unique refcode, usually an ascending numeric sequence (e.g. 00000001) of 8 characters.

4.3 Correcting Database Entries

- Database entries can easily be corrected using PreQuest. Simply select the relevant database file, e.g. *mybase.ind* from the **Open** menu, and load. You then select and modify entries in the same way that you would for any other input file format. Each time that a different entry is selected you are prompted to save any changes already made.

5 INPUT FILE FORMATS

- The current list of input file formats should satisfy the needs of most users, however more file formats can be added if there is sufficient demand for them. Please notify the CCDC if you have any further suggestions on this matter.

5.1 CIF

- The Crystallographic Information File is documented fully by Hall, Allen & Brown (*Acta Cryst.*, 1991, **A47**, 655-685).
- CIF is the preferred input file format as it is a globally accepted standard format and most of the popular crystallographic structure solution packages now provide a CIF output option.
- Not all of the CIF fields have to be provided, but in general it is best to provide as much data as is available. Do not include non-bonded or intermolecular distances in the CIF, as these cannot be stored in the CSD and will result in much 3D editing in PreQuest.
- A concordance list of CIF fields that can be read by PreQuest is given below. As the program converts CIF data to the BCCAB format for ease of editing within PreQuest, the list is presented as BCCAB headings with the CIF data fields that pertain to them listed below:

REFCODE (see page 34)
#JRNL (see page 35)
#AUTH (see page 35)
#DISORD (see page 35)
#QUAL (see page 35)
#PROPS (see page 36)
#CELL (see page 37)
#DENSITY (see page 38)
#SYSCAT (see page 38)
#RFACT (see page 38)
#UNIS (see page 38)
#CLASS (see page 38)
#COMPND & #SYNONM (see page 38)
#ATOM (see page 38)
#BOND (see page 39)
#FORMUL (see page 39)
#RMARKS (see page 39)

REFCODE

- From `_database_code_CSD` if present.
- If not then it will use `_journal_page_first` otherwise created from position of CIF file in

total list.

#JRNL

- From `_journal_coden_Cambridge`, `_journal_volume`, `_journal_page_first`, `_journal_year`
- If `_journal_coden_Cambridge` is not present then the program attempts to translates to the correct coden from `_journal_coden_ASTM` if present
- If a journal is one that starts from 1 with each issue then `_journal_issue` is also used.

#AUTH

- The only data name used is `_publ_author_name`

#DISORD

- Nothing done at the current time.

#QUAL

- "`POWDER:` powder data are output if any of the following fields are present and contain valid data:
  - `_pd_proc_ls_prof_R_factor`
  - `_pd_proc_ls_prof_wR_factor`
  - `_pd_proc_ls_prof_wR_expected`
  - `_pd_block_id`
  - `_pd_phase_block_id`
  - `_pd_calc_method`
- The last field is output to *INFO* if present.

- "`RADIATION: neutron` or "`RADIATION: synchrotron` are output if one of the fields:
  - `_diffrn_radiation_probe`
  - `_diffrn_radiation_type`
  - `_diffrn_radiation_source`
- contain `eutron` or `ynchrotron` respectively (the initial letter is ignored to avoid problems with capitalisation).

- The measurement temperaure "`TEMP:` is obtained from:
  - `_diffrn_ambient_temperature`
  - `_cell_measurement_temperature`

- The former takes precedence. A warning is given if only the latter is available, or where the values disagree by more than 10K. Where the measurement unit is indicated as deg.C in the file, this is converted to K for internal comparison, but output in the same units as in the input file. If a negative temperature, assumed to be in K, is detected it is only output in *INFO*. Any e.s.d. of the temperature is discarded.

- `"PRESSURE: at kPa` is obtained from the field:
  - `diffrn_ambient_pressure`
- This is assumed to be in kPa, and is output with these units.

- `"REINT-OF:` is obtained from the field
  - `_citation_database_id_CSD`

- The following are output to *INFO* (`CIFER REINT-OF:`) if present:
  - `_citation_journal_full`
  - `_citation_journal_volume`
  - `_citation_page_first`
  - `_citation_year`
  - `_citation_special_details`

- Other **#QUAL** subkeywords are output as a blank template, as follows:
  - `"ABS-CONFIG:`
  - `"POLYMORPH:`
  - `"BIOACTIVITY:`
  - `"CONFORMER:`
  - `"ISOMER:`
  - `"STEREOMER:`
  - `"RACEMATE:`
  - `"REINT-SEE:`
  - `"REFINEMENT:`
  - `"CONTRIB:`
  - `"OTHER-EXP:`
  - `"OTHER-DIFF:`

#PROPS

- The following fields are examined

  - _chemical_melting point
  - _exptl_crystal_colour _pd_char_colour
  - _exptl_crystal_preparation
  - _exptl_crystal_description _pd_char_particle_morphology
- and output in a template of the form

```
#PROPS    "MP.
          "COLOR:
          "HABIT:
          "SOURCE-CHEM:
          "SOURCE-NAT:
          "SENS:
          "ISO:
          "OTHER-CELL:
          "PHASE-T:
```

- These placeholders are output even if the relevant data are not located in the file.
- The melting point is output in deg.K, and converted from deg.C if this is indicated in the file. A warning is generated if the melting point is < 293K, and if now melting point is found. > is converted to above in the output, and : to -. Any e.s.d. on the melting point is ignored.
- If the colour is given as *transparent* or *clear*, a warning is given and nothing is written to the color field. The spelling of *colourless* is Americanised.
- The crystal source and habit are taken from the preparation and description fields respectively, where present.

#CELL

- Main cell data is obtained from:

```
_cell_length_a
_cell_length_b
_cell_length_c
_cell_angle_alpha
_cell_angle_beta
_cell_angle_gamma
_symmetry_cell_setting
```

- Z value from _cell_formula_units_Z
- Space group from _symmetry_space_group_name_H-M
- Cell volume from _cell_volume

#DENSITY

- Measure density no longer archived.
- Diffraction density from `_exptl_crystal_density_diffrn`
- Chemical formula weight from `_chemical_formula_weight`

#SYSCAT

- The crystal system is determined during *#CELL* processing.

#RFACT

- The lowest (non-zero) value of the following is used for #RFACT:

```
_refine_ls_R_factor_all
_refine_ls_R_factor_obs
_refine_ls_wR_factor_all
_refine_ls_wR_factor_obs
```

#UNIS

- Currently only the "template" is included in the BCCAB file.

#CLASS

- Currently only the stub "#CLASS" is output.

#COMPND & #SYNONM

- The following datanames are examined `_chemical_name_systematic`, `_chemical_name_common`.
- If both are present then `_chemical_name_systematic` goes into *#COMPND* and `_chemical_name_common` goes into *#SYNONM*.
- If only one is present this is put in *#COMPND*. If the only one is `_chemical_name_common` then a warning is put into *#RMARKS*.
- If neither is present then the title `_publ_section_title` if present, is used.
- If this is not present then `_publ_section_abstract` is used.

#ATOM

- Atom data is obtained from:

```
_atom_site_label
_atom_site_fract_x
```

```
_atom_site_fract_y
_atom_site_fract_z
```

#BOND

- Bond data is obtained from:

```
_geom_bond_atom_site_label_1
_geom_bond_atom_site_label_2
_geom_bond_distance
_geom_bond_site_symmetry_1
_geom_bond_site_symmetry_2
```

#FORMUL

- Formula data is obtained from `_chemical_formula_moiety`
- and if not present from `_chemical_formula_sum`
- If this is not present then from `_chemical_formula_iucr`
- and if this is not present then from `_chemical_formula_structural`
- *#CONN*, *#CREF*, *#CRYCON*, *#DIAGRAM*, *#RADIUS* and *#SYMM*
- Nothing done at the current time.

#RMARKS

- `"CCDC:` is output if the `field _database_code_depnum_ccdc_archive` (CCDC deposition number) is present. CCDC is prepended in not already present.
- `"XREF`: blank field is always output.
- `"REMARK:` blank field is always output.

#INFO

- Text after `data_` is written to *INFO* unless this is general or global. In addition to the messages generated in the above processing, the following fields are searched for the keywords listed below.
  - `_publ_section_title`
  - `_publ_section_abstract`
  - `_publ_section_comment`
  - `_publ_section_discussion`
  - `_publ_section_experimental`
  - `_publ_section_exptl_prep`
  - `_publ_section_exptl_refinement`
  - `_refine_special_details`

- `_exptl_special_details`
- `_exptl_crystal_preparation`
- `_chemical_compound_source`
- `_refine_ls_hydrogen_treatment`
- `_pd_prep_conditions`
- `_pd_proc_ls_special_details`
- `_pd_char_special_details`

- The current list of keywords is as follows

disorder-	occupanc-	mp	M.p.
Mp.	isostructural	isomorphous	Deuter-
guest	host	inclusion	clathrate
hydroscopic	moisture	decompose	absolute
optical	Friedel	anomalous	dispersion
Bijvoet	Hamilton	eta	Roger
R-factor ratio	Flack	activit-	inhibitor
agent	drug	medic-	-agonist
biological	bacteri-	microbial	viral
pharmac-	therapeut-	polymorph-	modification-
recrystalli-	crystallis-	crystalliz-	agostic
SQUEEZE-	Sluis	twin	crystal source
Reitveld			

- The sentence containing the keyword is written to *INFO*, provided that there is sufficient space. Note that there is a order of precedence for the keywords, which influences the order in which the sentences appear in *INFO*.
- Any non-unitary values of `_atom_site_occupancy` are identified up to a preset limit.

5.2 SHELX

- This format has been in existence since 1976, with some variations. The PreQuest program will cope with most files from different versions of SHELX. You may provide the SHELX instruction file if that is all that is available (*filename.ins*) but it is much better to provide the

output file which contains the coordinate e.s.ds, and the list of bond lengths with e.s.d. (*filename.res*).

- Since the SHELX format does not include the space group symbol PreQuest uses the *LATT* and *SYMM* fields to construct the pattern of operators stored in CSD, and looks up a definitive table of space group symbol versus symmetry operators to provide the text symbol e.g. Pbca.
- No special handling of disordered atoms is carried out, these will require attention using the 3D Edit function.

5.3 MOL

- PreQuest accepts MOL files: the input file for the MDL MACCS system. MOL files (Dalby *et al.*, *J. Chem. Inf. Comput. Sci.*, **1992**, 32, 244-255) are widely used for storage of Cartesian coordinate data, e.g. from NMR studies.
- PreQuest extracts the atom and bond blocks, preserving the MDL bond types (single, double, and triple bonds only). Any leading information in lines 1 to 3 of a MOL file will be treated as the *Compound* title.
- **Note:** In order for the CSD and ConQuest to operate uniformly over Cartesian and fractional crystallographic coordinates we use a convention whereby the coordinate data are placed in a cubic unit cell of axes 100Å, the space group is set as P_1, and the coordinates are stored as x/100, y/100, z/100.
- All intramolecular 3D searches work on the Cartesian coordinate entries in the same way as they do for ordinary main file crystallographic entries.

5.4 MOL2 Format

- This format is used by the Tripos Sybyl program and many molecular modelling "structures" are stored using it. PreQuest extracts as much information as can be stored in the CSD format and the first line following <*TRIPOS*> *MOLECULE* is copied to *#COMPND* as the compound name.
- Note that Tripos atom types cannot be stored in CSD, only the element symbol. The <*TRIPOS*> *BOND* fields are stored with bond types as given.

5.5 BCCAB Format

- This is the format used in-house at the CCDC for the main-file database creation. It is also the internal format used in PreQuest; each entry is stored in a temporary working file using this format. When an entry is edited in PreQuest using a text-editor this BCCAB format will appear, regardless of the original input format (SHELX etc.). The description given here is sufficient to allow the user to achieve two objectives:

1. to edit with the text-editor

2. to define the important fields for a stand-alone input record

- An entry in BCCAB is defined as a set of text lines of maximum length 80 characters. The first line must begin with the "**?**" character, followed by the reference code. The last line is *#END*.
- For private databases it is best to use numeric reference codes which will not be confused with main-file records. The reference code must be the single unique identifier for each record, e.g. 00001234. This is the key reference number used in the CSD database system.
- A typical input record for a published structure is shown below:

```
?HEMKOZ01
#JRNL 35,59,2787,1994
#AUTHOR R.Gleiter,B.Treptow,H.Irngartinger,T.Oeser
#CELL a 15.284(2) b 14.882(2) c 7.805(1) z 4 sg Pnma cent 1
#FORMUL C20 H26 O4
#COMPND Dimethyl
hexacyclo(7.7.0.0$2,8!.0$2,16!.0$8,10!.0$10,16!)hexadeca-1,9-
dicarboxylate
#SYNONM Dimethyl propella(3)prismane dicarboxylate
#DENSITY dx 1.240
#RFACT R= 0.0540
#PROPS "COLOR: colorless
"MP:
"HABIT:
"SOURCE-CHEM:
"SOURCE-NAT:
"SENS:
"ISO:
"OTHER-CELL:
"PHASE-T:
#QUAL "TEMP: at 243 K
"PRESSURE:
"RADIATION:
"POWDER:
"REFINEMENT:
"ABS-CONFIG:
"POLYMORPH:
"BIOACTIVITY:
"CONFORMER:
"ISOMER:
"STEREOMER:
"RACEMATE:
"REINT-SEE:
```

```
"REINT-OF:
"CONTRIB:
"OTHER-EXP:
"OTHER-DIFF:
#ATOM O1 0.8519,1 0.4546,1 0.0761,2
O2 0.7778,1 0.3475,1 -0.0662,2
C1 0.8856,1 0.3022,1 0.1253,3
C2 0.9865,1 0.3024,1 0.1573,3
C3 0.9251,1 0.3025,1 0.3078,3
C4 0.9150,1 0.3661,1 0.4540,3
C5 0.9732,2 0.3380,2 0.6058,3
C6 0.9476,3 0.2500 0.6921,5
C7 1.0560,1 0.3662,2 0.1005,3
C8 1.1455,1 0.3382,2 0.1703,4
C9 1.1831,2 0.2500 0.0997,6
C10 0.8376,1 0.3760,1 0.0449,3
C11 0.7316,2 0.4181,2 -0.1575,4
#BOND O1 C10 1.214,3
O2 C10 1.329,3
O2 C11 1.453,3
C1 C1* 1.554,3
C1 C2 1.562,3
C1 C3 1.548,3
C1 C10 1.462,3
C2 C2* 1.560,3
C2 C3 1.504,3
C2 C7 1.492,3
C3 C3* 1.562,3
C3 C4 1.491,3
C4 C5 1.539,4
C5 C6 1.525,3
C7 C8 1.530,3
C8 C9 1.535,3
#END
```

- The entry consists of a number of data fields. Each field begins on a new line with the character "#" followed by the field name. There are no restrictions on the order of fields, or the spacing within the text. The suggested minimum requirement for a private database is the following:

  ?refcode
  #COMPND : compound name
  #AUTHOR : person responsible for data

#CELL : unit cell and space group
#ATOM : atomic coordinates
#END

- It is strongly recommended that as much data as possible is entered at the time of input as it will enrich the database and make the record much more useful to future users. You should check through the fields listed below and provide data if you can. Remember that *#QUAL* and *#RMARKS* text is "searchable" by ConQuest, and you can use your own keywords here.

A Note on Chemical Diagrams

At the CCDC a 2D chemical diagram is constructed using the **2D Edit** function of PreQuest, which will appear as the two fields *#CONN* and *#DIAG*. There is no need to describe the format of these fields here - if they appear in a working record, please do not edit them. In general you do not input the diagram by typing these fields. *PreQuest* will make the diagram either automatically (see Make 2D), or using the graphical interface (see 2D Edit).

#ATOM : Atomic information (see page 44)
#AUTHOR : Authors' names (see page 45)
#BOND : Bond lengths (see page 45)
#CELL : Unit Cell Data (see page 46)
#CLASS : Chemical Class (see page 47)
#COMPND : Compound Name (see page 48)
#FORMUL : Molecular Formula (see page 48)
#JRNL : Journal Reference (see page 49)
#PROPS : Crystal Properties (see page 49)
#QUAL : Qualifier (see page 50)
#RADIUS : Elemental Radius for bonding (see page 51)
#RFACT : R-Factor (crystallographic refinement accuracy) (see page 52)
#RMARKS : Remarks (see page 52)
#SYNONM : Compound Synonym (see page 53)
#SYSCAT : Crystal System and Entry Category (see page 53)
#TOLER : Bonding Tolerance (see page 54)
#UNIS : Processing Information (see page 54)

#ATOM : Atomic information

- Each atom has an atom label, and fractional coords x/a, y/b, z/c with optional estimated standard deviations (e.s.d.). e.g.

```
#ATOM C1   0.1234,3  -0.3456,12  0.4567,8
C1'        0.3456,2   0.2345,13  0.3456,7
```

Atom labels

- Maximum length 8 characters. Must begin with a valid element symbol, usually followed with numbers, e.g. Br2, C123, but any string of alphabetic characters or quote mark is allowed, e.g. C1' C1" H11a Ow1

Atom coordinates

- Input these with a decimal point. If an e.s.d. is given type this after the comma and with no space. e.g. 0.1234,12 means e.s.d. is 0.0012.

Suppressed flag

- Suppressed atoms have an "S" after the z-coordinate. These can be manually written using the text editor. In the example below C1' is suppressed.

```
#ATOM C1   0.1234,3  -0.3456,12  0.4567,8
C1'        0.3456,2   0.2345,13  0.3456,7  S
```

#AUTHOR : Authors' names

- Author names to be written in the following style:

```
#AUTHOR A.B.Smith, J.-P.Mornon, P.Van Stappen, G.L'Abbe,
A.Fforbes-Hamilton, D. van der Helm, Yu.T.Struchkov,
R.King III, E.F.Meyer Junior, Shao Mei-cheng
```

Note

Give initials with full-stop, no spaces, and use comma to separate names. This is consistent with the main *CSD* file and enables concurrent searching.

#BOND : Bond lengths

- This field contains the bond lengths reported by the author and corresponds to the atomic coordinates listed in #ATOM. Each bond length is described using the appropriate pair of atom labels followed by the value of the distance. If the e.s.d. of the bond length is available then it follows the value and is separated from it by comma. E.g.

```
#BOND  C1 C12  1.451,3
C1 H1  0.98,1
```

Note

This is optional input. These author-given bonds are used as a consistency check in **Check-3D**, comparing calculated and given values and are not vital in *PreQuest*.

#CELL : Unit Cell Data

- This field contains all of the unit cell information using a variety of keywords:
  - **a**: length of unit cell a-axis (Å)
  - **b**: length of unit cell b-axis (Å)
  - **c**: length of unit cell c-axis (Å)
  - **alpha**: value of interaxial angle alpha (in degrees)
  - **beta**: value of interaxial angle beta (degrees)
  - **gamma**: value of interaxial angle gamma (degrees)
  - **v**: volume of unit cell (in cubic Angstroms)
  - **z**: number of formula units per unit cell
  - **sg**: space group symbol (in Hermann-Mauguin notation)
  - **cent**: flag to indicate whether or not the space group origin is at a centre of symmetry

- An example of a *#CELL* field is:

```
#CELL    a 6.3746,5   b 15.8638,8   c 7.7460,6
         alpha 87.12,1   beta 91.34,4   gamma 93.67,4   v 776.42
         z 4   sg P-1   cent 1
```

- This example of an anorthic cell can be used to illustrate various details:
  - Standard deviations of unit cell parameters given after comma
  - Interaxial angles are expressed in decimal form (not degrees/minutes)
  - Only the minimum set of unit cell parameters is recorded:

```
Anorthic      a  b  c  alpha  beta  gamma
Monoclinic    a  b  c         beta
Orthorhombic  a  b  c
Tetragonal    a     c
Rhombohedral  a        alpha
Hexagonal     a     c                gamma
Cubic         a
```

- In most reported studies the monoclinic cell is chosen with the b-axis unique so beta is recorded. However, if the a-axis is unique then alpha is recorded and likewise if the c-axis is unique then gamma is recorded.
  - The unit cell volume is expressed as a real number. If the author reports v = 1234 it is recorded as v 1234 (v is used only for check purposes and is not archived to the database).

- The z value is recorded as an integer, (number of formula units per cell).
- The space group symbol is recorded in Hermann-Mauguin notation with the use of two conventions. The bar symbol above a symmetry axis symbol is replaced by the sign in front of the axis symbol. Suffixes which are normally involved in screw axis symbols are recorded "in-line" e.g. 21 for a 2-fold screw axis.

- Special conventions are used for the recording of monoclinic space group symbols for a- or c-axis unique:

```
a-axis unique      P21 is recorded as P2111
                   P21/n is recorded as P21/n11

c-axis unique      P21 is recorded as P1121
                   P21/n is recorded as P1121/n
```

- If a trigonal space group is described in terms of a rhombohedral unit cell (a and alpha recorded) then the conventional space group symbol, e.g. R3, is recorded as R3r.
- The cent flag is program-generated from the space group symbol and takes the values 1 or 2.
  - **cent 1** indicates that the space group origin is at a centre of symmetry
  - **cent 2** indicates that the space group origin is not at a centre of symmetry
- The cent flag is directly linked to the set of general equivalent positions (*#SYMM* field) which is program-generated from the space group symbol. For cent 1 the *#SYMM* field contains only one half of the general equivalent positions - those not related by the centre of symmetry at the origin.
- Certain space groups allow a choice of origin and the program default always chooses the setting with a centre of symmetry at the space group origin. If this choice is incorrect for a particular structure determination then cent 2 should be manually set and the appropriate *#SYMM* field input manually.

#CLASS : Chemical Class

- This field contains the chemical class assignment for the compound. These classes are listed below. Example:

```
#CLASS   1 8
```

Note

Each entry can be assigned up to 2 class numbers. Steroids are not classified as terpenes. A single space is used to separate classes. This is optional data but can be very useful especially for classifying natural products.

Class Number	Class of Compounds
1	Carbohydrates and their metal complexes
2	Nucleosides, nucleotides and their metal complexes
3	Amino-acids, peptides and their metal complexes
4	Porphyrins, corrins and their metal complexes
6	Steriods
7	Terpenes
8	Alkaloids
11	Organic polymers

#COMPND : Compound Name

- This field contains the chemical compound name following standard rules if possible. For natural products the trivial name is usually recorded - a field exists to supplement the name by one or more synonyms e.g. drug or trade name (see *#SYNONM*).
- CSD conventions are:

1. Greek lower-case letters are spelt out in full e.g. alpha
2. Subscripts and superscripts are indicated by the use of `!sub$`, and `$super!`
3. Metal oxidation states are recorded e.g. iron(0), cobalt(ii), molybdenum(vi)

Note
You can use trivial names or local names e.g. Compound A12387

#FORMUL : Molecular Formula

- This field contains the molecular formula, represented as the sum of the individual formulae for each of the residues. A residue is defined here as being a discrete bonded unit. For example, sodium acetate monohydrate consists of 3 residues *viz*. acetate anion, sodium cation, water molecule - its formula is recorded as:

```
#FORMUL  C2 H3 O2 1-,Na1 1+,H2 O1
```

- The general expression for a residue formula is:

  pre-multiplier (elements with atom counts charge) *post-multiplier*

- In polymeric structures, all residues are given the multiplier `n` (or `2n`, `3n` etc) even if they are not, in themselves, polymeric.

- Elements are listed in the order:

  C followed by H followed by D
  followed by other elements in alphabetic order

Note
This is always used at CCDC for consistency checking. If you use PreQuest to construct the 2D chemical diagram (*Make-2D* or *Edit 2D*) the *#FORMUL* field is generated automatically. A *#FORMUL* field is necessary for screen generation by PreQuest and without it the entry may be rendered unsearchable in the database.

In the example:

```
#FORMUL  C2 H3 O2 1-,Na1 1+,H2 O1
```

notice that spaces separate each element-item, commas separate residues. Charges are given at end of residue in the style 1-, 2+ etc.

#JRNL : Journal Reference

- This field contains the journal reference for a published structure. It takes the form: coden, volume, page, year. E.g.

```
#JRNL  591,39,136,1983
#JRNL  1078,,,1995
```

Note
Optional input for private databases. If #JRNL is not input PreQuest will treat the records as *Private Communication, coden 1078*, and fill in the current year from the computer system date. A coden is a table of code number pointing to a text name for a journal.

#PROPS : Crystal Properties

- This field stores crystal property information using different keywords

  "`MP`: the effect of heating on a crystal
  "`COLOR`: the colour of the crystal at room temperature in daylight
  "`HABIT`: the appearance of the crystal
  "`SOURCE-CHEM`: the recrystallisation solvent(s) used to prepare the crystal, or any other method used to obtain the crystal (e.g. sublimation; from the melt etc.)
  "`SOURCE-NAT`: the natural source of the chemical compound
  "`SENS`: the rapid instability of the crystal under normal storage conditions
  "`ISO`: referal to isostructural, isomorphous and isotypic analogues

"OTHER-CELL: other cell parameters for the same compound/polymorph or a twin compound
"PHASE-T: the temperature(s), pressure or other conditions of the phase transition(s)

- Example:

```
"MP:   293deg.C.
"COLOR:  red
"HABIT:  prism
"SOURCE-CHEM:  dichloromethane/hexane
"SOURCE-NAT:  bark of Japoni
"SENS: air-sensitive
"ISO: isostructural with the 2-thio analogue
"OTHER-CELL:  cell parameters of a second crystal were a= 9.99,2
b= 8.28,7 c= 12.67,3
"PHASE-T:  undergoes a first-order phase transition, to the
rhombohedral polymoprh, at 426 K
```

Note
Optional input but potentially contains valuable information - include if relevant data are available.

#QUAL : Qualifier

- This field contains important attributes of the compound, or the crystallographic study if different from X-rays at room temperature.
- The types of data recorded in main-file *CSD* are:
  - "TEMP: the temperature at which the x-ray experiment was carried out; input either as C (displays deg.C.) or if no units input, displays K.
  - "PRESSURE: the pressure at which the x-ray experiment was carried out; blank field implies atmospheric pressure.
  - "RADIATION: type of radiation used in the experiment; blank field implies x-ray.
  - "POWDER: powder study; y is inserted if the experiment was carried out on a powder sample.
  - "REFINEMENT: any non-standard details of the refinement process.
  - "ABS-CONFIG: y inserted if the absolute configuration has been determined by x-ray methods.
  - "POLYMORPH: polymorph descriptor.
  - "BIOACTIVITY: details of any biological activity, or potential biological activity of the compound, or a closely related analogue.
  - "CONFORMER: descriptor of conformation.
  - "ISOMER: descriptor of isomer.

- "STEREOMER: insert refcode of any stereoisomers of the compound under study.
- "RACEMATE: insert refcode of the racemate of the compound under study.
- "REINT-OF: insert refcode of the original structure, if the current study uses the same data but reassigns the space group and/or connectivity.
- "REINT-SEE: insert refcode of the reinterpreted structure, if the current data has been subsequently reinterpreted with a change to the space group and/or connectivity.
- "CONTRIB: details of the source of the data, when not from a published article.
- "OTHER-EXP: any other details regarding the compound under study.
- "OTHER-DIFF: any other details of the diffraction experiment.

Example:

```
"TEMP: 307
"PRESSURE: 1.2 kbar
"RADIATION: neutron
"POWDER: y
"REFINEMENT: high-order
"ABS-CONFIG: y
"POLYMORPH: monoclinic
"BIOACTIVITY: potential anticancer
"CONFORMER: 1,3-alternate
"ISOMER: high-melting
"STEREOMER: SACFAZ
"RACEMATE: WIMZIM WUMZOM
"REINT-OF: YEYNAR
"REINT-SEE: YEYNAR01
"CONTRIB: contribution from the thesis of T. Plesnivy,
Inst.Org.Chem.,Univ.Mainz,Germany (1995)
"OTHER-EXP: irradiated product
OTHER-DIFF: thermal decomposition study; non-merohedral twin
```

Note
Optional for private databases - but recommended input.

#RADIUS : Elemental Radius for bonding

- This field contains the radius used to determine the crystal connectivity for each element present in the list of atomic coordinates (**#ATOM**). The distance D*ij* between two atoms *i* and *j* is defined to be a bonding distance if

(R*i* + R*j* - Tol) .le. D*ij* .le. (R*i* + R*j* + Tol)

where R*i*, R*j* are radii of atoms and *Tol* is a tolerance normally 0.4 Å.

Note
Not normally input. The radius values are obtained from a standard table by PreQuest. The user will not normally need to edit these values as radius adjustment can be performed by Radj in Check 3D. The table is printed in the Quest documentation, *Vol 3. Appendix 10*.

#RFACT : R-Factor (crystallographic refinement accuracy)

- This field contains the crystallographic R-factor as a decimal number e.g.

```
#RFACT R=0.0410
```

- The sub-keyword R= has the value 0.0410, indicating an R-factor of 4.1%

Note
Optional input - but a very valuable indicator of accuracy.

#RMARKS : Remarks

- This field is used to record deposition numbers, cross references to identical structures and editorial comment resulting from CCDC validation, details of unresolved problems including those resulting from program limitations.
- Text comments are separated by semi-colon(;) which should NOT be used within the text comment.
- The sub-fields are:
  - "CCDC: CCDC deposition number.
  - "XREF: cross-reference to identical structures (with identical data).
  - "REMARK: remarks regarding the compound or structure not recorded elsewhere.

- Example

```
"CCDC: CCDC 103518
"XREF: REDFAN displays as

"XREF: Coordinates may be obtained from REDFAN

as in the CSD, coordinates would be deleted from the subsequent
publication
```

```
"REMARK: 40% void space in the structure; crystallizes as a
conglomerate; Mo-Mo bond order is 3.5; large structure in which
program limits have been exceeded.
```

Note
Optional input - but this is searchable text in ConQuest.

#SYNONM : Compound Synonym

- This field contains any appropriate synonym(s) for the compound name. It is often used to record trade or drug names. If more than one synonym is required then separate with semi-colon. Examples:

```
#SYNONM  Aspirin
    #SYNONM   Ampicillin; Nuvapen; Totapen
    #SYNONM   1,8-Dihydroxy-2,4,5,7-tetranitro-9,10-
anthracenedione
```

Note
Optional input. Local company names could be given e.g. A1234C5.

#SYSCAT : Crystal System and Entry Category

- This field contains the crystal system information using the keywords **sys** and **cat** e.g. #SYSCAT sys A cat
- Crystal system is recorded as:

  sys A for anorthic (triclinic)
  sys M for monoclinic
  sys O for orthorhombic
  sys T for tetragonal
  sys H for hexagonal
  sys R for rhombohedral
  sys C for cubic
- For trigonal space groups the conventions used are:
  - If space group lattice symbol is R, e.g. R-3, we set **sys R**
  - If space group lattice symbol is P, e.g. P31, we set **sys H**
- Category is always recorded as 3, meaning full structure determination.

Note
Not normally input - generated automatically by PreQuest.

#TOLER : Bonding Tolerance

- This field contains the tolerance, in Angstroms, used in conjunction with **#RADIUS** for the determination of the crystal connectivity. See #RADIUS. E.g.

```
#TOLER 0.40
```

Note

Not normally input - generated automatic by PreQuest.

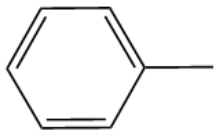
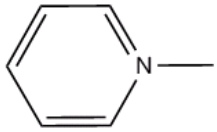
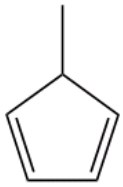
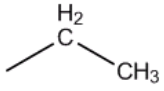
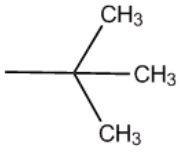
#UNIS : Processing Information

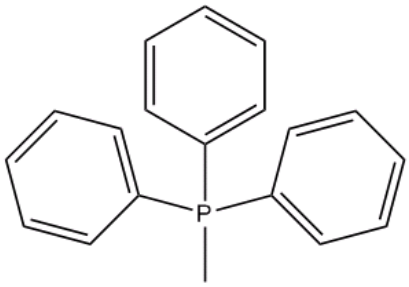
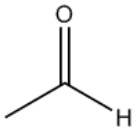
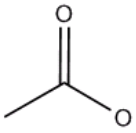
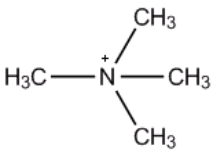
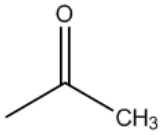
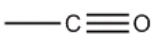
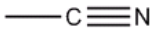
- This field is used to control the level of processing of the data, causing certain test to be by-passed. It is chiefly for use by the CCDC editorial staff to allow unresolved errors to be flagged in the main database, and to ensure that no incorrect connectivity is stored. The data consists of keywords followed by values.

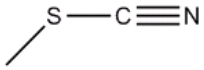
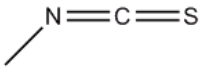
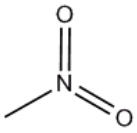
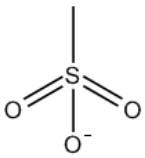
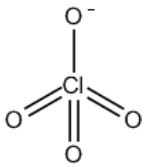
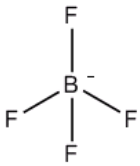
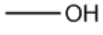
```
int  3  diffractometer data
     2  densitometer photographic data
     1  visual photographic data

ig   1  ignore atom valence checks

rpa  1  refer problem to author - by letter

pd   1  disorder is present (set by PreQuest if the #DISORD
"COMMENT: field contains text).
```

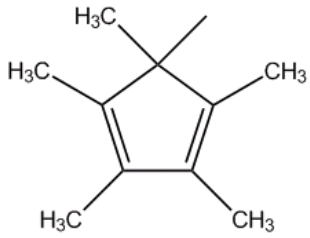
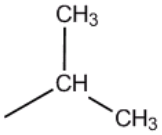
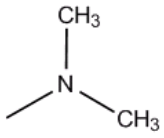
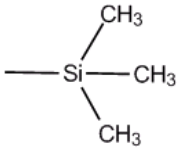
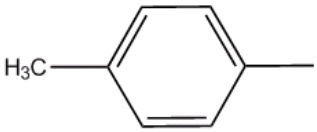
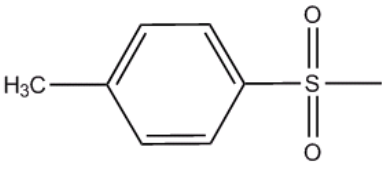
Appendix 1: Group Symbols

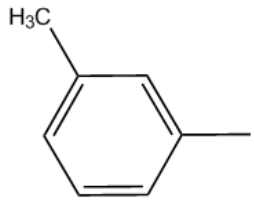
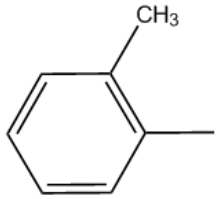
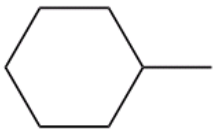
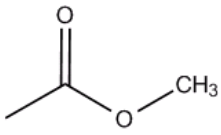
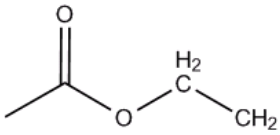
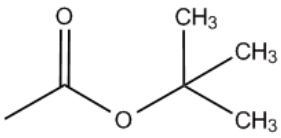
Note that Keyboard Shortcuts are case insensitive.

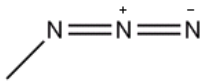
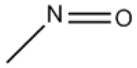
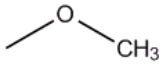
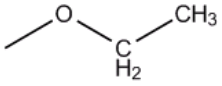
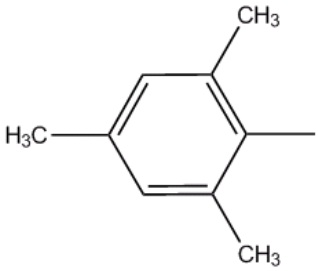
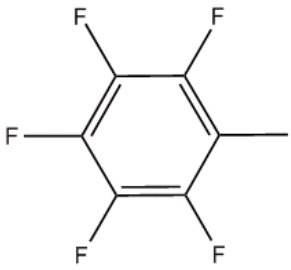
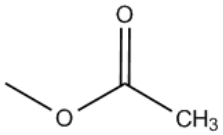
Name	2D connectivity of representation	2D Display	Keyboard Shortcut
Phenyl		Ph	PH
Pyridyl		Py	PY
Cyclopentadienyl		Cp	CP
Ethyl		Et	ET
t-Butyl		Bu^t	BU
Trifluoromethyl		CF_3	TF

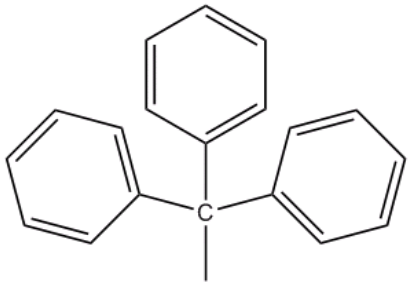
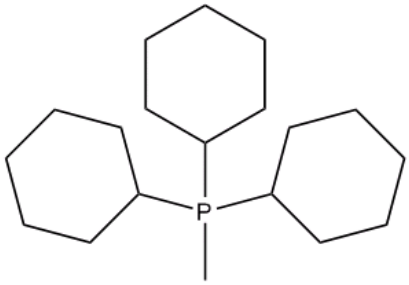
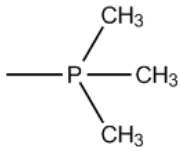
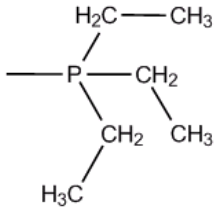
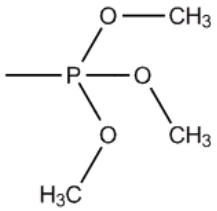
Triphenylphosphine		PPh_3	TP
Formyl		CHO	AY
Carboxylate		COO	XA
Tetramethyl ammonium		Me_4N^+	MA
Acetyl		COMe	MK
Carbonyl		CO	CY
Cyano		CN	CN

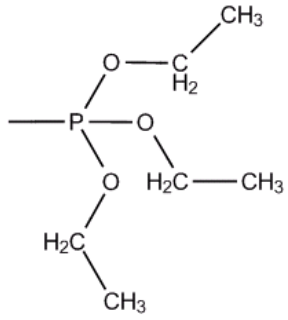
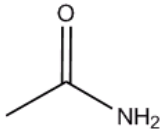
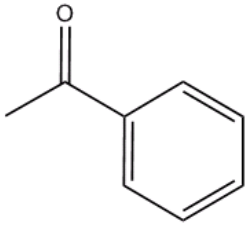
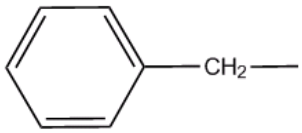
Thiocyanato		SCN	TO
Isothiocyanato		NCS	IO
Nitro		NO_2	NT
Sulfonate		SO_3^-	SU
Perchlorate		ClO_4^-	PC
Tetrafluorborate		BF_4^-	FB
Hydroxyl		OH	OH

Pentamethylcyclopen-tadienyl		Cp*	MP
i-Propyl		Pr^i	IP
Dimethylamino		NMe_2	NM
Trimethylsilyl		$SiMe_3$	TS
p-Tolyl		pTol	PL
p-Tosyl		pTos	PS

m-Tolyl		mTol	MT
o-Tolyl		oTol	OT
Cyclohexyl		Cy	CX
Methoxy carbonyl		COOMe	MC
Ethoxy carbonyl		COOEt	EC
t-Butoxy carbonyl		$COOBu^t$	BC

Azide		N_3	AZ
Nitroso		NO	NS
Methoxy		OMe	OM
Ethoxy		OEt	OE
Mesityl		Mes	MS
Fluorobenzyl		C_6F_5	FP
Acetoxy		OCOMe	AK

Triphenylmethyl		CPh_3	TY
Tricyclohexylphos-phine		PCy_3	HX
Trimethylphosphine		PMe_3	PX
Triethylphosphine		PEt_3	PE
Trimethylphosphite		$POMe_3$	MX

Triethylphosphite		$POEt_3$	EX
Acetamide		$CONH_2$	CB
Benzoyl		COPh	BZ
Benzyl		CH_2Ph	BN